THE FORMATIVE YEARS OF RELATIVITY

A cause for celebration. In September 2015, gravitational waves, generated by a collision of two black holes, were detected by the twin Laser Interferometer Gravitational-wave Observatory (LIGO) detectors, operated by Caltech and MIT, confirming a consequence of the general theory of relativity established by Albert Einstein 100 years earlier. Drawing by Laurent Taudin.

THE FORMATIVE YEARS OF RELATIVITY

THE HISTORY AND MEANING OF EINSTEIN'S PRINCETON LECTURES

Featuring Einstein's Classic Text
The Meaning of Relativity
in Its Historical Context

HANOCH GUTFREUND
AND JÜRGEN RENN

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Copyright © 2017 by Princeton University Press
Requests for permission to reproduce material from this work should
be sent to Permissions, Princeton University Press
Published by Princeton University Press, 41 William Street, Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press, 6 Oxford Street,
Woodstock, Oxfordshire OX20 1TR
press.princeton.edu
Jacket photograph from "Professor Einstein's Visit to the United States," *The Scientific Monthly,* Vol. 12, No. 5 (May 1921)
All Rights Reserved

Library of Congress Cataloging-in-Publication Data

Names: Gutfreund, Hanoch, author. | Renn, Jürgen, 1956- author.
Title: The formative years of general relativity : the history and meaning of Einstein's Princeton lectures / Hanoch Gutfreund and Jürgen Renn.
Description: Princeton ; Oxford : Princeton University Press, [2017] | Includes bibliographical references and index.
Identifiers: LCCN 2016047532| ISBN 9780691174631 (hardcover ; alk. paper) | ISBN 0691174636 (hardcover ; alk. paper)
Subjects: LCSH: General relativity (Physics)—History—20th century. | Einstein, Albert, 1879–1955. Relativitäatstheorie. English.
Classification: LCC QC173.6 .G878 2017 | DDC 530.11—dc23 LC record available at https://lccn.loc.gov/2016047532

British Library Cataloging-in-Publication Data is available
This book has been composed in New Century Schoolbook, Minion Pro
Printed on acid-free paper. ∞
Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

TURNING TO THE THEORY OF RELATIVITY ITSELF I AM ANXIOUS TO DRAW ATTENTION TO THE FACT THAT THIS THEORY IS NOT SPECULATIVE IN ORIGIN; IT OWES ITS INVENTION ENTIRELY TO THE DESIRE TO MAKE PHYSICAL THEORY FIT OBSERVED FACTS AS WELL AS POSSIBLE. WE HAVE HERE NO REVOLUTIONARY ACT BUT THE NATURAL CONTINUATION OF A LINE THAT CAN BE TRACED THROUGH CENTURIES. THE ABANDONMENT OF CERTAIN NOTIONS CONNECTED WITH SPACE, TIME, AND MOTION HITHERTO TREATED AS FUNDAMENTALS MUST NOT BE REGARDED AS ARBITRARY, BUT ONLY AS CONDITIONED BY OBSERVED FACTS.

Einstein, "On the Theory of Relativity," Lecture
at King's College, London, 1921

THE THEORY OF RELATIVITY IS A FINE EXAMPLE OF THE FUNDAMENTAL CHARACTER OF THE MODERN DEVELOPMENT OF THEORETICAL SCIENCE. THE INITIAL HYPOTHESES BECOME STEADILY MORE ABSTRACT AND REMOTE FROM EXPERIENCE. ON THE OTHER HAND, IT GETS NEARER TO THE GRAND AIM OF ALL SCIENCE, WHICH IS TO COVER THE GREATEST POSSIBLE NUMBER OF EMPIRICAL FACTS BY LOGICAL DEDUCTION FROM THE SMALLEST POSSIBLE NUMBER OF HYPOTHESES AND AXIOMS. MEANWHILE THE TRAIN OF THOUGHT LEADING FROM THE AXIOMS TO THE EMPIRICAL FACTS OR VERIFIABLE CONSEQUENCES GETS STEADILY LONGER AND MORE SUBTLE. THE THEORETICAL SCIENTIST IS COMPELLED IN AN INCREASING DEGREE TO BE GUIDED BY PURELY MATHEMATICAL, FORMAL CONSIDERATIONS IN HIS SEARCH FOR A THEORY, BECAUSE THE PHYSICAL EXPERIENCE OF THE EXPERIMENTER CANNOT LEAD HIM UP TO THE REGIONS OF HIGHEST ABSTRACTION.

Einstein, "The Problem of Space, Ether and the
Field in Physics," *My Worldview*, 1934

CONTENTS

Foreword
DIANA KORMOS BUCHWALD ix

Preface xiii

PART I. PRELIMINARIES 1

1. Introduction 3

2. Einstein's First Trip to America 11

3. Structure and Contents of The Meaning of Relativity 17

PART II. THE EMERGING WORLD OF GENERAL RELATIVITY 23

1. Physics and Geometry 25

2. The Principles of General Relativity 34

3. The First Solutions and the Challenge of Their Interpretation 46

4. Einstein and the Astronomers 52

5. The Genesis of Relativistic Cosmology 69

6. The Controversy over Gravitational Waves 94

7. Philosophical Debates on General Relativity 106

8. The Quest for a Unified Field Theory 122

9. Early Monographs on Relativity 140

10. Beyond the Formative Years 155

PART III. EINSTEIN'S BOOK WITH THE APPENDIXES 159

1. Space and Time in Pre-relativity Physics 161

2. The Theory of Special Relativity 184

3. The General Theory of Relativity 215

4. The General Theory of Relativity (Continued) 239

5. Appendix for the Second Edition: On the "Cosmologic Problem" 269

6. Appendix II (PUP 4th edition, 1953): Generalization of Gravitation Theory 293

7. Appendix II (PUP 5th edition, 1956): Relativistic Theory of the Non-symmetric Field 326

Part IV. The Popular Lectures 361

1. Introductory Remarks 363

2. Einstein's Lectures 366

Part V. Biographical Notes on the Protagonists of the Formative Years 387

Index 409

FOREWORD

Almost a century has elapsed since Albert Einstein visited the United States for the first time in the spring of 1921, six years after having published a remarkable series of papers that signaled the completion of his general theory of relativity, the most significant development in theoretical physics since Isaac Newton's *Principia* more than two centuries earlier.

Einstein extemporaneously delivered five lectures at Princeton University during the first week of May 1921;[1] six hundred invitations were sent out for these Stafford Little Lectures.[2] A large number of guests, among them close to three hundred visiting scientists, came to town, and more than four hundred attended Einstein's inaugural lecture in McCosh 50. Princeton University's president John G. Hibben (1861–1933) hired a stenographer for the purpose of recording Einstein's lectures. But over the next few days, Einstein's audiences dropped significantly: on the second day the auditorium was "half-filled. On the third day the discussion was moved to a small classroom. And as the explanation progressed through its one hour and a half-length, in groups of two and three the scientists dropped out of the discussion and out of the room."[3] But each lecture received exhaustive coverage in newspapers.

On his first day in town, a two-page memorandum of agreement was drawn up with Princeton University Press to the effect that "Albert Einstein being the Author and Proprietor of a work entitled Lectures delivered at Princeton University May 9–13, 1921, in consideration of the covenants and stipulations hereinafter contained, and agreed to be performed by the Publishers, grants and guarantees to said Publishers and their successors the exclusive right to publish the said work in all forms during the terms of copyright and renewals." The Press agreed to pay Einstein 15% of "their Trade-List (retail) price" for all cloth copies sold, no earlier than "six months after the date of publication." Einstein would be entitled to six free copies of the work. "If, on the expiration of five (5) years from date of publication, or at any time thereafter, the demand for said work should not, in the opinion of said Publishers be sufficient to render its publication profitable, then upon written notice by said Publishers to said Author this contract shall cease." The contract was signed by Einstein and the manager of the Press, Paul G. Tomlinson (1888–1977).

The publishers' clause would never be invoked, as demand for the book, which came to be known as *The Meaning of Relativity*, never flagged.

* * *

Einstein spent an entire workweek in Princeton. He gave his first lecture on Monday, 9 May, after having been presented earlier in the day with an honorary doctoral degree.

While in Princeton, Einstein learned that Dayton C. Miller of Case Institute, Cleveland, had found a non-zero ether drift during preliminary experiments performed at Mount Wilson Observatory in April. He apparently responded with what was to become a famous phrase, "Raffiniert ist der Herrgott, aber boshaft ist er nicht" (Subtle is the Lord, but malicious he is not), to be inscribed a decade later above the fireplace in the common room of the newly built Fine Hall, home to Princeton's mathematics department, albeit in a rather unpoetic and incorrect translation: "God is clever, but not dishonest."

After visiting Princeton, Einstein spent two more weeks touring the East Coast and then sailed for England, where he stopped in Manchester and London. He laid a wreath at Newton's grave in Westminster Abbey, visited Oxford, and returned to Berlin in mid-June 1921. He began working on the seventy-three-page manuscript of the Princeton lectures in early September. He wrote to his close friend Maurice Solovine (1875–1958), who had been eager to embark on a French translation of the lectures, that he was working feverishly on the manuscript ("Ich schreibe auf Tod und Leben am Manuskript") but was making slow progress owing to some "wretched sluggishness."[4]

Two likely early drafts of the Princeton manuscript are available. The first, entitled "On the Special and General Theory of Relativity," is a nine-page document containing brief treatments of his first, third, and fourth lectures. In it, Einstein for the first time explicitly stated that the field equations of general relativity imply the equations of motions for particles in a gravitational field. But he did not include such a statement in the published version of the Princeton lectures and wouldn't do so until five years later, in a paper coauthored with Jakob Grommer in 1927.[5] The second draft, dealing with the special theory of relativity, was an attempt to re-create only his first Princeton lecture.[6]

At the end of September 1921, Tomlinson wrote from Princeton, inquiring for a second time as to when the Press might expect the manuscript. They were "considerably alarmed at the long delay in the receipt of your manuscript, which we were led to believe would be in our hands within a month after the lectures were delivered."[7]

Einstein wrote on other topics while the Press was waiting. "On the Misery of Children," written on 7 October 1921, was a testimonial for the US relief organization Central Committee for the Relief of Distress in Germany and Austria, and expressed publicly what Einstein had said privately to friends already two years earlier, namely, shock at the extent of urban want and child mortality at home. In mid-October, Einstein left Berlin for more than six weeks to lecture and visit in Italy, Switzerland, and Holland. He returned in late November and completed an unrelated yet important paper on a possible experiment to decide between a wave theory and a quantum theory of light.[8] Further, he wrote a brief article on the impact of science on the development of pacifism and another for the Austrian newspaper *Neue Freie Presse* on the plight of German science at a time of grave economic difficulties. He also hosted several events in support of the establishment of the Hebrew University.

He must therefore have returned to his Princeton lectures sometime late in 1921. He completed the manuscript in early January and sent it to the German publisher Friedrich Vieweg & Sohn on 9 January 1922, asking that it be typeset "as soon as possible." He requested that space be reserved for marginal comments, and also that several sets of page proofs be sent to him so that he could forward some to his English translator. He indicated that the German edition could appear only after the book's publication in America. "I insist

on the compensation of 20% of the sales price because not only the paper and printing costs but also everything that the publishing house needs for its existence and the author needs for living has risen proportionately and is still rising." Vieweg had offered 10%. A slight embarrassment ensued when Einstein realized that he had earlier promised the manuscript to the German publisher Springer Verlag as well. He also planned to send Solovine the typed proofs, to be translated by him into French and published by Gauthier-Villlars, on condition that they await the US publication.[9]

On 20 February 1922, Einstein was able to dispatch a letter and the German language page proofs to Oswald Veblen (1880–1960), professor of mathematics at Princeton. He asked that they be transmitted for translation to Edwin P. Adams (1878–1956), professor of physics at Princeton, who had served as the simultaneous translator of Einstein's oral lectures. Alas, as Veblen reported on 6 April, the manuscript arrived "considerably later than your letter, but I finally received it and turned it over to Professor Adams. The Princeton University Press is taking the steps necessary for its publication." Indeed, Tomlinson confirmed only a few days later that the English translation was complete. He also wrote that "it has been suggested by one of the members of the Mathematics Department here at Princeton that possibly the book would have more appeal to the general reader if certain passages in it which are rather technical could be explained and somewhat simplified. The suggested method of doing this is either by a Prefatory article or an explanatory appendix. Before proceeding, however, we wish to have your approval, and I should appreciate it greatly if you would write me whether you have any objection to this or not. We feel that it might help the sale of the book." It is possible that Veblen himself had made the suggestion. Einstein had no objections to a possible explanatory preface or appendix written by someone else, provided that the text of the manuscript remained unchanged. He sent a summary that could be used as a postscript, but that summary is now lost. It may be that the summary was used as the preface published in the German edition. By late May 1922, the German publisher Vieweg expressed impatience and surprise that the US edition had not yet appeared.[10]

Adams must have worked very diligently on the translation, because by July 1922 the British publisher Methuen, who had secured the rights for a UK publication of the book, deemed the English translation excellent and was hoping for successful sales. By the fall of 1922, a Polish translation of the German text was already in the making.[11]

An important contributor to this edition of Einstein's book was his stepdaughter, Ilse Einstein, who not only handled most of the correspondence surrounding the publication process of its several editions but apparently also devoted herself to the content of the book. We know that on 20 November 1922 she sent the corrected proofs of the German manuscript to Vieweg, while Einstein was away on his tour of the Far East. Ten days earlier Einstein had been awarded the Nobel Prize in Physics for 1921.

By December 1922, the final proofs of the German edition had been printed and sent to Ilse Einstein for review. Copies were sent out to Austria and Spain.[12] The German publisher planned for a first edition of three thousand copies.[13] A second edition was printed soon thereafter.

Documents reveal that, at the same time, Princeton University Press and Methuen collaborated during all book production steps, including its printing. An index card at Princeton University Press gives precise information: the book, the 126th volume in the Press's eighteen-year history, with its dark blue cloth cover and a gold pencil sketch on the front cover, gold-flat backbones, no headband, no gilt top, and trimmed all around, carried the publication date

of 12 January 1923, while its printing by Aberdeen University Press and its copyright were registered as 1922. The same book had been produced for Methuen by the same printers and must have been released only a few weeks earlier in the UK. Princeton University Press obtained 2,560 copies, with tan wraps printed in black. It spent $608.85 on printing costs, $486.40 on binding, and $14.47 for dies, for a total cost of $1,335.99 plus $9 for cut. The total cost per copy amounted to 53 cents; the book was to be sold for $2 per copy, the rough equivalent of about $30 in today's value. The stenographer's fee and expenses of $128.85 were noted, as was the fact that the stenographic record was never used in the production of the book. The University President's Report for the year 1922, published by Princeton University Press in the *Alumni Weekly*, noted the book's appearance. For the Methuen UK edition, Einstein was offered some $6 per copy sold and an advance of $3,000 in in today's currency.[14]

* * *

The story of the book does not end here, as Hanoch Gutfreund and Jürgen Renn beautifully explain in the present volume. It was revised, amended, and corrected by Einstein throughout the next three decades of his life and has now been reproduced in countless printings and languages throughout the world.

The present new edition is testimony to the enduring importance of Einstein's book, still considered by many to be the best exposition of general relativity. *The Meaning of Relativity* has been in print continuously, through many editions. Gutfreund and Renn have rendered an invaluable service by providing the contemporary reader with an updated and annotated edition that places the book and its revisions in the context of recent historical scholarship. Their most significant contribution with this volume is to show, and explain, how the *Princeton Lectures* came to represent a "significantly more advanced stage in the development of general relativity, in the understanding of its implications, and in the reception of the theory by the scientific community."

—Diana Kormos Buchwald

NOTES

1. "A Conversation with Einstein," interview in the Berlin newspaper *Vorwärts*, 30 June 1921, CPAE vol. 12, Appendix F, p. 540.
2. CPAE vol. 12, Introduction, p. xxxviii.
3. *Evening Bulletin* (Philadelphia), 12 May 1921, p. 20, quoted in Jozsef Illy, ed., *Albert Meets America* (Baltimore: Johns Hopkins University Press, 2006), p. 194.
4. AEA 21-151, draft letter to Maurice Solovine in Ilse Einstein's hand, CPAE vol. 12, Calendar, entry of 7 September 1921. The French edition was published in November 1924 by Gauthier Villars.
5. CPAE vol. 7, Doc. 63 and its note 6, and CPAE vol.15, Doc. 443.
6. CPAE vol. 7, Doc. 64.
7. CPAE vol. 12, Doc. 251.
8. See "On an Experiment concerning the Elementary Process of Light Emmision," *Königlich Preussische Akademie der Wissenschaften. Sitzungsberichte* (1921): 882–883, CPAE vol. 7, Doc. 68.
9. CPAE vol. 13, Doc. 16, 14 January 1922.
10. CPAE vol. 13, Calendar, Abs. 250.
11. CPAE vol. 13, Calendar, Abs. 397.
12. CPAE vol. 13, Calendar, Abs. 475.
13. CPAE vol. 13, Calendar, Abs. 105.
14. CPAE vol. 12, Calendar entry for 22 June 1921, p. 459.

PREFACE

This is our third book on the history and contexts of the early days of Einstein's theory of general relativity and his new gravitational physics. As in the other books,[1] we feature a canonical text by Einstein around which we build the narrative. This book focuses on *The Meaning of Relativity*, which is based on the lectures Einstein delivered in Princeton, New Jersey, in May 1921, during his first trip to the United States.

The Meaning of Relativity has been reprinted in many editions and translated into many languages, but rarely with any historical commentaries. Although this is a classical text, it does not reflect the full scope of what was known about general relativity and its consequences at that time because Einstein's intention was to expose the fundamentals in the thought process of the theory as clearly as possible. What it therefore does reflect is a particular stage in the development of Einstein's theory and of his own thinking about this development and his personal attitude toward the ongoing discussions. What it also reflects is a particular moment in the evolution of modern physics that has rarely been the focus of historical investigations.

This book discusses the debates and developments characterizing the early reception and spread of Einstein ideas in the late 1910s and 1920s. We have called this period the formative years of general relativity. It is one of the greatest transformation processes in the history of science because within a few years these ideas changed from the esoteric pursuit of an individual into not only the concerns of a growing community of scientists and philosophers but also a matter of international public and media attention.

It is only natural that the manifold activities of this period have attracted the attention of both historians and philosophers of science. This has led to numerous studies of the different aspects of the formative years, as they have appeared, for instance, in the volumes of the *Einstein Studies*.[2] This process of investigation has been significantly fostered by the preparation and publication of the relevant volumes of the *Collected Papers of Albert Einstein*,[3] from which we have also benefited in our work. In many cases, we cite the English translation of the original documents.

Nevertheless, so far no comprehensive account of the formative years has appeared that is accessible also to nonspecialists. This is what we hope to achieve with this book. Whenever possible, we refer to the original sources by Einstein and his peers. For readers interested in technical details or original historical investigations, references to the relevant literature are given.

We are grateful to Al Bertrand from Princeton University Press for guiding us through the different stages of this project and to the referees appointed by PUP for their extremely

valuable suggestions. Special thanks are due to our colleagues and friends Alexander Blum, Yemima Ben-Menachem, Jean Eisenstaedt, Fynn Ole Engler, Michel Janssen, David Kaiser, Roberto Lalli, Cormac O'Raifeartaigh, David E. Rowe, Donald Salisbury, and Robert Schulmann. Their remarks have been enormously useful in finalizing the manuscript.

We are grateful to the staff of the Albert Einstein Archives at the Hebrew University for their assistance, specifically to its director, Roni Grosz, to Barbara Wolff, to Chaya Becker, and to Orith Burla. We would like to acknowledge the professional and effective assistance in bibliographical research of Sabine Bertram. We also thank Florian Abe for his assistance with the images.

This book contains the first English translation of the two popular Princeton lectures presented to a broader nonprofessional audience. We are grateful to Dieter Brill for translating and editing them.

We would also like to warmly thank Diana Kormos Buchwald, the general editor of the Einstein Papers Project, for her continuing support of our endeavors and for her hospitality in Pasadena, where part of this book was written. Our visit there coincided with the discovery of gravitational waves, which added to the excitement of working on the origins of the new gravitational physics. There we also benefitted from interactions and discussions with members of her staff, specifically with Dennis Lehmkuhl.

One of us (H. G.) is grateful to the Max Planck Institute for the History of Science for the hospitality extended during numerous visits in the course of this work. One of us (J. R.) appreciates the hospitality of the California Institute of Technology during the period as Bacon Visiting Professor in 2016, which coincided with the visit at the Einstein Papers Project.

This project is in special debt to two institutions that were directly and indirectly involved in the project: the Hebrew University, which allowed us to use archival material, and the Max Planck Institute for the History of Science where this project was mostly pursued. We are therefore grateful for the support of these two institutions.

Finally, we acknowledge with appreciation and gratitude the invaluable editorial assistance and professional support by Lindy Divarci from the shaping of the idea until the completion of the published work.

NOTES

1. Hanoch Gutfreund and Jürgen Renn, *The Road to Relativity: The History and Meaning of Einstein's "The Foundation of General Relativity" Featuring the Original Manuscript of Einstein's Masterpiece* (Princeton, NJ: Princeton University Press, 2015); Hanoch Gutfreund and Jürgen Renn, eds., *Relativity: The Special and the General Theory, 100th Anniversary Edition* (Princeton, NJ: Princeton University Press, 2015).
2. Don Howard and John Stachel, eds., *Einstein Studies*, published under the sponsorship of the Center for Einstein Studies, Boston University (New York: Springer (Birkhäuser), 1989–).
3. *The Collected Papers of Albert Einstein* (CPAE; Princeton, NJ: Princeton University Press, 1987–), http://einsteinpapers.press.princeton.edu/.

PART I

PRELIMINARIES

1
INTRODUCTION

The Meaning of Relativity, also known as *Four Lectures on Relativity*, is Einstein's definitive exposition of his special and general theories of relativity. It was written in the early 1920s, a few years after he had elaborated his general theory of relativity. Neither before nor afterward did he offer a similarly comprehensive exposition that included not only the theory's technical apparatus but also detailed explanations making his achievement accessible to readers with a certain mathematical knowledge but no prior familiarity with relativity theory. In 1916, he published a review paper that provided the first condensed overview of the theory but still reflected many features of the tortured pathway by which he had arrived at his new theory of gravitation in late 1915. An edition of the manuscript of this paper with introductions and detailed commentaries on the discussion of its historical contexts can be found in *The Road to Relativity*.[1] Immediately afterward, Einstein wrote a nontechnical popular account, *Relativity—The Special and General Theory*.[2] Beginning with its first German edition, in 1917, it became a global bestseller and marked the first triumph of relativity theory as a broad cultural phenomenon. We have recently republished this book with extensive commentaries and historical contexts that document its global success. These early accounts, however, were able to present the theory only in its infancy.

Immediately after its publication on 25 November 1915, Einstein's theory of general relativity was taken up, elaborated, and controversially discussed by his colleagues, who included physicists, mathematicians, astronomers, and philosophers. Einstein himself also made further fundamental contributions to the development of his theory, exploring consequences such as gravitational waves and cosmological solutions, elucidating concepts such as that of the energy and momentum of the gravitational field, and even reinterpreting basic aspects of the theory. A turning point was the confirmation of the bending of light in a gravitational field, which, as predicted by general relativity, was observed during a solar eclipse in 1919. These were the *formative years* of relativity in which the theory essentially received the structure in which it later became one of the pillars of modern physics. *The Meaning of Relativity* is the paradigmatic text of this period, reflecting not only Einstein's own efforts but also the engagement of his contemporaries with the theory. Einstein evidently returned to the theory of relativity in many later publications, both specialized and popular. He later also enriched *The Meaning of Relativity* with appendixes discussing further developments. But he never made another attempt at such an all-encompassing presentation in which he painstakingly motivated, explained, and discussed its basic principles and their consequences.

In this book, we place *The Meaning of Relativity* in the context of the formative years of general relativity, which extended from 1915 until around the blossoming of quantum mechanics in the late 1920s and early 1930s, when more and more leading physicists turned their attention to this booming new field. This is the period when the basic ideas and principles of relativity theory crystallized and were refined, when the first solutions of the field equations were derived and analyzed, and when the cosmological consequences of the theory were debated. All of these events happened in the context of an expanding albeit fragile network of scientists who, at least for a certain time, became interested in the theory and enriched it with their contributions: from Hendrik Antoon Lorentz to Erwin Schrödinger, from Karl Schwarzschild to Willem de Sitter, from David Hilbert to Hermann Weyl, from Moritz Schlick to Hans Reichenbach. Einstein's text not only reflects the considerable maturation that the theory of general relativity had undergone since its creation, it also represents a key document of the scientific sociodynamics characterizing its formative years.

As a characteristic stage in the development of the theory, these formative years have, remarkably, received less attention from historians than subsequent periods, such as the "low-water-mark period" (Jean Eisenstaedt) immediately following it,[3] in which, also due to the Second World War, the fragile network mentioned above was partly destroyed and the general theory of relativity was perceived, with the exception of its cosmological implications, as a rather esoteric subject outside of mainstream physics. A "renaissance" (Clifford Will) of relativity would only come more than a decade after the war and in connection with new astrophysical discoveries in the 1960s, such as those of quasars.[4] Actually, as has recently been argued, the renaissance of general relativity had begun even earlier, essentially as the result of a community-building effort turning the theory into a universally applicable framework.[5] This revival was followed by what has been called the "golden age" (Kip Thorne) of relativity,[6] which witnessed new conceptual insights, such as those into the nature of spacetime singularities, and turned the theory into the foundation of modern astrophysics and observational cosmology.

Still, the advances made in the formative years were considerable. These years saw the further pursuit of the conceptual revolution that Einstein had started in 1907, when he first introduced the equivalence principle and began his search for a comprehensive theory describing both gravitation and inertia within a relativistic framework. At the end of 1915, he had found the field equations of this theory, but their meaning and implications were still far from being obvious. Scientific revolutions are hardly the result of a sudden paradigm shift in the sense characterized by the historian of science Thomas Kuhn.[7] The gravitational field theory published in November 1915 was the result of a laborious process that involved integrating the classical knowledge about gravitation embodied in Newton's theory with the relativistic concepts of space and time established by Einstein in 1905 in the context of what later came to be known as special relativity.

The profound conceptual and far-reaching technical consequences of the theory were, in this first moment, still unexplored territory. Even the extent to which the very principles that had formed Einstein's heuristic starting point were actually realized by the definitive formulation of the theory was not clear. The formative years thus saw—and this is one of the hallmarks of this period—a fundamental revision and reinterpretation of just these basic principles, including the relativity principle itself. In this process of revision, clarification, and reinterpretation, the theory became increasingly independent from the elements of classical physics that Einstein's impressive edifice had used as its scaffolding.[8]

More and more, general relativity became an autonomous conceptual framework in its own right. This process was not completed by the end of its formative years and still continues today. With the publication of *The Meaning of Relativity*, however, a turning point was reached, because Einstein's masterful exposition made it possible to study and understand the theory in its own right.

The Meaning of Relativity is based on the five lectures delivered by Einstein as the Stafford Little Lectures at Princeton University over five consecutive days, starting on 9 May 1921. The first two lectures, attended by large audiences, were popular accounts of the special and the general theory of relativity. The other three lectures were given to smaller audiences of scientists. The lectures were delivered in German. It was agreed from the outset that the lectures would be published according to the following plan:

> A German stenographer is taking notes of the lectures as they are delivered. The plan of procedure is to have her write her notes out in German and then Professor Edwin P. Adams of the Department of Physics will go over them and check up those scientific portions which may have caused trouble. After Professor Adams has completed this part of the work, the lectures will be submitted to Professor Einstein for revision and final approval. When he has returned them, they will be translated into English and published.[9]

Each lecture was summarized orally in English by Professor Adams. These summaries were published on the days following each lecture in the *New York Evening Post* (the first four lectures) and in the *New York Times*. The extensive press coverage of these lectures demonstrates the great public interest generated by Einstein's visit to Princeton.

Between the beginning of September 1921 and early January 1922, Einstein summarized the Princeton lectures for publication. He produced a manuscript entitled "Four Lectures on the Theory of Relativity," which he submitted to his German publisher, Vieweg.[10] The manuscript is divided into five lectures, but in the published version the first two are combined into one. The first German edition was published in 1922, followed soon after by a second edition in 1923. The English translation by Edwin P. Adams, entitled *The Meaning of Relativity*, was published in London by Methuen in 1922 and then by Princeton University Press in 1923. Subsequent English editions, in which Einstein still had an input, appeared in 1945, 1950, 1953, and 1956.

Einstein's introduction to the first German edition, which was not reproduced in the English editions, reads:

> In the present elaboration of the four lectures I gave at Princeton University in May 1921, I wanted to summarize the principal thoughts and mathematical methods of relativity theory. In this I have made an effort to leave out everything that is less essential, but to treat the fundamentals in such a way that the whole may serve as an introduction for all those who master the elements of higher mathematics, but do not want to spend all too much time and effort on the subject. This short treatment evidently makes no pretense to completeness, in particular as I have not dealt with the more subtle, more mathematically interesting developments founded in variational calculus. My principal aim was to let the fundamentals in the entire train of thought of the theory emerge clearly.[11]

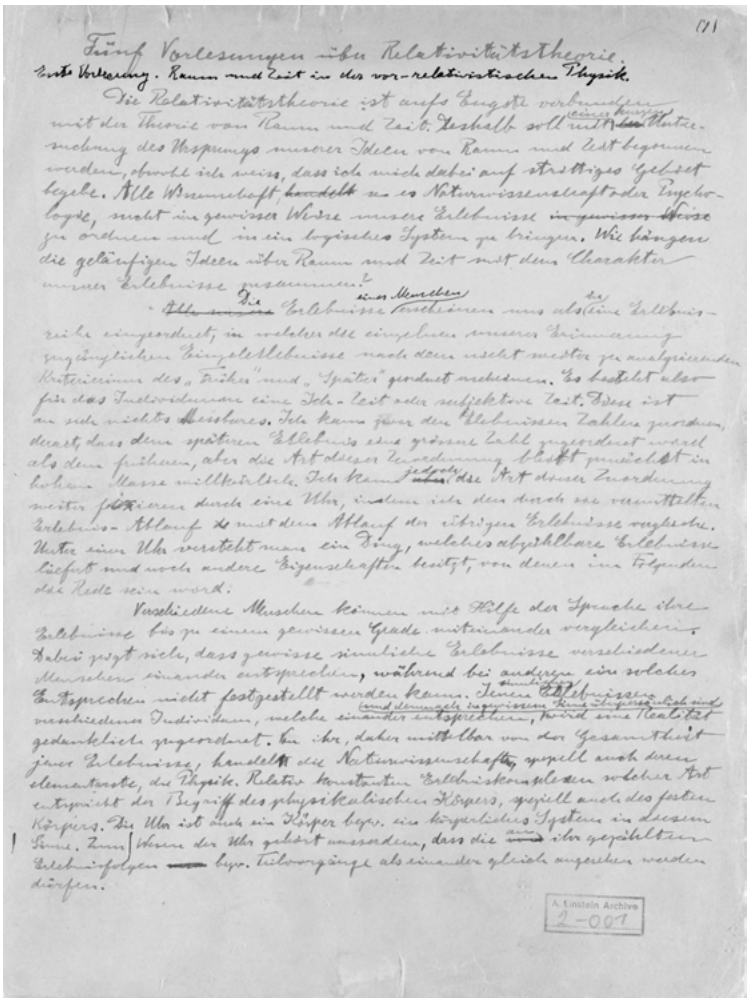

Fünf Vorlesungen über Relativitätstheorie.
Erste Vorlesung. Raum und Zeit in der vor-relativistischen Physik.
A. Einstein Archive
2-001

FIGURE 1: The first page of the manuscript *Five Lectures on the Theory of Relativity*.

As mentioned above, in 1917 Einstein had published his book *Relativity—The Special and the General Theory (A Popular Account)*, through which he intended to make the basic ideas of the theory accessible to a general audience. Its introduction begins with a concise description of the book's aim:

> The present book is intended, as far as possible, to give an exact insight into the theory of relativity to those readers who, from a general scientific and philosophical point of view, are interested in the theory, but who are not conversant with the mathematical apparatus of theoretical physics.[12]

In his Princeton lectures, in contrast, Einstein intended to present these ideas in a coherent fashion to the scientists "who master the elements of higher mathematics."[13]

In our view, however, this is not the most important difference between these two canonical texts. What is more important is that there is a five-year interval between them. During these years, in extensive correspondence with colleagues and in several publications on the fundamental questions and ideas of the theory of relativity, Einstein responded to critical remarks and participated in debates about the physical and cosmological implications of general relativity. As a consequence, *The Meaning of Relativity* includes new formulations of the basic principles and ideas and a change of emphasis on specific issues. It represents a significantly more advanced stage in the development of general relativity, in the understanding of its implications, and in the reception of the theory by the scientific community.

In this book, we trace the roots of these developments in the contemporary research of Einstein and others, and at the same time we comment on how these developments are reflected in Einstein's text, with the aim of making it accessible to modern readers in a new way, as the key document of these formative years.

The book is divided into four parts. Following this introduction, chapter 2 in part I sets the historical context for Einstein's visit to Princeton in May 1921. Chapter 3 of this part discusses the structure and content of his book. Then, in part II, we offer a detailed portrait of the formative years of general relativity. In part III, Einstein's text is reproduced together with the appendixes.

Each of the chapters of part II analyzes a specific dimension of this historical period and at the same time serves as a reading aid to Einstein's book, describing the relevant developments in his thinking and how they are reflected in his text. Our commentaries do not aim to explain Einstein's text, which has remained one of the most lucid expositions of his theory to this day. We instead focus on those issues that will set *The Meaning of Relativity* into the historical context of what we have called the formative years. Occasionally, we also point to advances that were still to come in later periods of the development of general relativity. While Einstein's text employs mathematical formulas to explain his theory, our historical commentaries are aimed at a wider audience and endeavor to present the relevant concepts and developments without using such formalism, making them more broadly accessible.

Chapter 1 of part II deals with the ongoing debate about the meaning of Einstein's theory in the formative years, with particular attention to the relation between physics and geometry. Einstein himself devoted major interpretational efforts to this theme, as in his famous paper "Geometry and Experience" or in his musings about the role of the ether concept in the theory of general relativity. This chapter also compares Einstein's thinking on this issue with that of the French mathematician and philosopher Henri Poincaré and deals with the role of symmetry in the theory of relativity, one of Einstein's enduring legacies. The role of symmetry becomes evident, for instance, in the lecture on special relativity, in which it is shown how relativistic invariance, a symmetry property of the spacetime continuum, shapes Maxwell's equations and other laws of physics. In the period under consideration, the understanding of symmetry is deepened by the emergence of Emmy Noether's famous theorems, for which the theory of general relativity was an important source of inspiration.

Chapter 2 of part II deals with the question of what the theory actually achieved and specifically reexamines the meaning of the relativity principle. This question of its meaning was raised by critical observers whose comments led to a partial reinterpretation of general

relativity. The German physicist Erich J. Kretschmann argued that the principle of general covariance has no physical content and only constitutes a mathematical requirement. This contention generated an exchange of letters in which Einstein conceded Kretschmann's criticism, but Einstein does not mention Kretschmann's remarks explicitly in his book. We shall discuss these developments and correlate them with his correspondence with colleagues and with other texts he published during the formative years.

The formative years also saw a first wave of the exploration of exact solutions to Einstein's gravitational field equations, beginning with the work of Karl Schwarzschild, Johannes Droste, Willem de Sitter, Alexander Friedmann, Hans Reissner, Gunnar Nordström, and Georges Lemaître, which is discussed in chapter 3.

Chapter 4 deals with the early exploration of observational and experimental consequences of general relativity, ranging from Erwin Freundlich's failed attempts to verify gravitational light bending and the redshift to the triumphal confirmation of light bending during a solar eclipse by Arthur Eddington's expedition, and finally to Hubble's discovery of the redshift of distant galaxies, which established the notion of an expanding universe.

Chapter 5 highlights the background for the emergence of relativistic cosmology. A specific interpretational issue, which is treated in this context, is the role of Ernst Mach's principle and the first exploration of the cosmological consequences of general relativity. Here Einstein's main interlocutor was the Dutch astronomer Willem de Sitter. But during the years 1916 to 1918, Einstein also exchanged letters with the mathematicians Hermann Weyl and Felix Klein. This correspondence has been referred to as the Einstein–de Sitter–Weyl–Klein debate. This debate not only focused on specific cosmological models but also helped to clarify the meaning of such fundamental issues as coordinates and energy-momentum conservation in the theory. The discussion of cosmology in the final Princeton lecture is influenced by this debate. This chapter also addresses the appendix "On the Cosmological Problem," which was added to the second English edition in 1945. It emphasizes the implications of Hubble's discovery.

Chapter 6 addresses one of the key themes of modern general relativity: gravitational waves. In 1916 Einstein performed the first relevant calculations, including a first derivation of the celebrated quadrupole formula, which, however, contained a mistake that he corrected in 1918. The corrected formula became the basis for the long-term search and the first observational verification of these waves. These calculations provoked a discussion about the reality of these waves, which continued well into the 1950s and beyond. This is the only major topic debated during the formative years that has no trace in *The Meaning of Relativity*. Had we restricted our commentaries to the contents of Einstein's book, there would be no reason to mention gravitational waves; however, it would be inconceivable to talk about the formative years without thoroughly discussing them. What is worth emphasizing in this context is how Einstein's predominant interest in this phenomenon, which developed immediately after the completion of his general theory, had faded away completely by the time he delivered the Princeton lectures.

Chapter 7 addresses the philosophical debate surrounding general relativity in its formative period, from Einstein's own writings on methodological and epistemological aspects of the theory to the first reactions by philosophers such as Moritz Schlick, Ernst Cassirer, and Hans Reichenbach. The focus is on the philosophical investigations clarifying and reinterpreting the conceptual foundations of general relativity.

In the formative years, starting with David Hilbert in 1915 and culminating in the works of Hermann Weyl, Theodor Kaluza, Oskar Klein, Arthur Eddington, and of course Einstein himself, general relativity was often seen as a partial or incomplete theory to be generalized, encompassed, or superseded by an overarching unifying theory that would also explain the constitution of matter, typically by including gravity and electromagnetism. Chapter 8 is dedicated to the contemporary debates of this issue and Einstein's role in it. Specifically, the chapter addresses the question of *why* general relativity was perceived to be an incomplete theory by many contemporary scientists, including Einstein himself. It will also discuss the second appendix, entitled "Generalization of Gravitation Theory," which Einstein added to the third edition of *The Meaning of Relativity*. This appendix was modified in subsequent editions, and its last version was entitled "Relativistic Theory of the Non-symmetric Field." It reflects Einstein's struggle in his search for a unified field theory, which in the last ten years of his life was based on the notion of a non-symmetric metric. We reproduce here the last two versions of this appendix, which are included in the fourth and the fifth editions.

Chapter 9 discusses the emergence of a tradition of writing textbooks on relativity and of teaching the subject. How did practitioners and students of relativity learn the theory and what were the different modes of access to it? Around 1920, in addition to Einstein's books and articles, a series of influential monographs and textbooks appeared, including those of Hermann Weyl, Wolfgang Pauli, Max Born, Max von Laue, and Arthur Eddington. We will compare these different presentations and analyze their influence on the later development of the theory. Alongside these monographs, which advanced the progress of scholarship in this new and expanding field, there was also an extensive anti-relativity literature. It can easily be discarded today as physical misunderstanding, as a trend serving the goals of rising nationalism in Germany, or as pure anti-Semitism. Nevertheless, that literature was part of the formative years and cannot be ignored in the present discussion as it also reflects on the broader social dynamics of the reception of the theory. Chapter 10 offers an outlook on the further transformations of Einstein's theory after the formative years.

Part IV of this book is dedicated to the first two popular lectures Einstein delivered in Princeton, which were not included in the published version. Chapter 1 offers an introductory commentary, while chapter 2 reproduces the extant text of these two lectures. These lectures have never before been published in English.

Because the formative years were a period in which the advancement of general relativity was furthered by Einstein's exchange with an ever-wider circle of physicists, astronomers, mathematicians, and philosophers who had become interested in his work, we will devote, in part V, biographical notes to some of Einstein's outstanding interlocutors of this period, among them Willem de Sitter, Felix Klein, Hendrik Lorentz, Hermann Weyl, Moritz Schlick, Hans Reichenbach, Ernst Cassirer, Gunnar Nordström, Gustav Mie, and Erich Kretschmann. Taken together, they illustrate that general relativity was maturing also in the sense that it was now becoming a community effort, even if this early community still remained fragile. The fervent debates and exchange of ideas are best reflected in the extensive correspondence between Einstein and its key players. To demonstrate the dynamics of these interactions, we frequently quote from this correspondence. In addition, we include as inserts or boxes in the main text brief explanations of basic concepts. We also reproduce historical documents such as newspaper articles or letters that illustrate the historical context of Einstein's Princeton lectures and their role as a pivotal moment in the history of modern physics.

NOTES

1. Gutfreund and Renn, *Road to Relativity*.
2. A centennial edition of this text with extensive commentaries is now available: Gutfreund and Renn, *Relativity: The Special and the General Theory*.
3. Jean Eisenstaedt, "The Low Water Mark of General Relativity, 1925–1955," in *Einstein and the History of General Relativity*, ed. D. Howard and J. Stachel (Boston: Birkhäuser, 1989), pp. 277–292.
4. Clifford Will, "The Renaissance of General Relativity," in *The New Physics,* ed. P. Davies (Cambridge: Cambridge University Press, 1989), pp. 7–33.
5. See Alexander Blum, Roberto Lalli, and Jürgen Renn, "The Reinvention of General Relativity: A Historiographical Framework for Assessing One Hundred Years of Curved Space-time," *Isis* 106 (2015): 598–620; and Alexander Blum, Roberto Lalli, and Jürgen Renn, "The Renaissance of General Relativity: How and Why It Happened," *Annalen der Physik* 528 (2016): 344–349; Alexander Blum, Domenico Giulini, Roberto Lalli, Jürgen Renn (eds.), "The Renaissance of Einstein's Theory of Gravitation," Special Issue *EPJ H,* 2017. See also Bernard F. Schutz, "Thoughts about a Conceptual Framework for Relativistic Gravity," in *Einstein and the Changing Worldviews of Physics*, ed. C. Lehner, J. Renn, and M. Schemmel (New York: Birkhäuser, 2012), pp. 259–269.
6. Kip S. Thorne, *Black Holes and Time Warps: Einstein's Outrageous Legacy* (New York: Norton, 1994), pp. 258–299.
7. Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago IL: University of Chicago Press, 1962).
8. See Michel Janssen and Jürgen Renn, "Arch and Scaffold: How Einstein Found His Field Equations," *Physics Today* 68 (November 2015): 30–36.
9. *Princeton Alumni Weekly*, 11 May 1921, pp. 713–714.
10. Albert Einstein, *Vier Vorlesungen über Relativitätstheorie* (Braunschweig: Vieweg, 1922).
11. CPAE vol. 7, p. 499 (German version) Translation by the authors.
12. Gutfreund and Renn, *Relativity: The Special and the General Theory*, p. 10.
13. See n. 11 above.

2
EINSTEIN'S FIRST TRIP TO AMERICA

Einstein became famous after the confirmation of his prediction of the bending of light from distant stars by the gravitational field of the sun. He subsequently received several invitations to lecture at American universities. In October 1920, Luther Eisenhart, a professor of mathematics at Princeton University, inquired on behalf of John Hibben, the president of the university and a professor of logic, if Einstein would agree to accept a special lectureship in the winter.[1] Einstein responded directly to Hibben, accepting the invitation with gratitude and stating that he would be able come to Princeton only in September 1921.[2] Following the advice of his friend and colleague, Paul Ehrenfest, who suggested that such a visiting lectureship should also be financially rewarding,[3] Einstein asked for a relatively high honorarium of $15,000, mentioning that he had made the same offer to another American university that had approached him (he meant the University of Wisconsin). Hibben responded that financial difficulties would make it impossible for Princeton to cover these costs and thus ended their correspondence.[4]

Things changed when Einstein agreed to join the Zionist delegation to the United States to promote plans for the establishment of a university that would later become the Hebrew University of Jerusalem. He wrote again to Hibben:

> I am compelled to travel to America at the invitation of the Zionist organization in the middle of March, because of the founding of the university to be established in Jerusalem. At this occasion I could hold a few lectures in Princeton, provided you deemed it desirable. . . . In addition to Princeton University, I also received invitations from the universities of Wisconsin and Yale and from the Academy in Washington. However, I address only you because your invitation was the first I received and because the time at my disposal is very limited."[5]

To appreciate how important this was to Einstein, one should mention that in order to make this trip he had declined an invitation to participate in the Solvay Conference that was to take place in April 1921. The Solvay Conferences constituted the most prestigious stage on which the great ideas and advances in physics were presented and debated. Einstein was the only German physicist to be invited to the first such conference after World War I. Yet when Kurt Blumenfeld, a leading figure in the German Zionist Organization, invited him

on behalf of Chaim Weizmann, head of the organization, to join the mission to the United States, he immediately agreed. In a letter to his friend the Dutch physicist Hendrik Antoon Lorentz,[6] Einstein thanked him for arranging the invitation to participate in the Solvay Conference and apologized for declining it because "the Zionists are planning to establish a university in Jerusalem and they believe that my trip to the US might influence rich Jews to contribute more generously. As strange as it sounds, they are probably right." He felt that it was his duty to help: "This initiative appeals to my heart and, as a Jew, I consider it my duty to contribute to its success as much as I can." Lorentz responded by expressing his understanding for Einstein's decision and wishing him success.[7] The prominent German–Jewish scientist Fritz Haber, Einstein's colleague and friend, was not so encouraging. He criticized Einstein's decision to go to the United States, invoking patriotic arguments in the name of his debt to German science and his loyalty to his German colleagues. He even used the argument that Einstein's decision might generate anti-Semitic sentiments against German Jews.[8] Einstein was undeterred. He explained that his decision stemmed from his obligation "to help my morally oppressed ancestral companions as much as is in my power," and he categorically rejected any allegations of disloyalty to his German colleagues.[9]

The group arrived in the United States on 1 April 1921. Together, they were to take a six-week, coast-to-coast tour to promote the cause of the Jewish national home and to raise funds for the establishment of Hebrew University in Jerusalem. The following day, the *New York Times* reported that thousands of people had waited four hours to welcome Professor Einstein to America.

On the group's arrival, Chaim Weizmann told reporters:

> Professor Einstein has done us the honor of accompanying us to America in the interest of the Hebrew University of Jerusalem. We have long cherished the hope of creating in Jerusalem a centre of learning in which the Hebrew genius shall find full self-expression and which shall play its part as interpreter between the Eastern and Western worlds.

To this, Einstein added: "I know of no public event which has given me such delight as the proposal to establish a Hebrew University in Jerusalem."[10] Indeed, on another occasion, Einstein referred to the establishment of the Hebrew University as "the greatest thing in Palestine since the destruction of the Second Temple."[11]

Einstein intended to spend most of his time in the United States promoting the cause of the university but was encouraged by Weizmann to devote part of his time to lecturing at American universities:

> It would be absolutely necessary for the success of our activities in America that you accept invitations from American universities to deliver scientific talks. . . . I naturally would not permit myself to influence you in any way in freedom of choice and would only like to convey to you the view of our friends in America that by accepting a few such invitations you would be supporting our work most effectively.[12]

This is the context in which Einstein's first visit to the United States and the Princeton lectures took place. The Princeton visit, like the rest of his tour of the United States, was covered in detailed day-to-day reports in the press.[13]

Einstein's first appearance before a scientific audience was when he delivered a lecture at Columbia University on 15 April 1921. Three days later he began a series of four lectures at the City College of New York (CCNY) entitled "The Special Theory of Relativity"; "The Development of the Special Theory"; "The General Theory of Relativity and Gravitation"; and "The Ether and Radiation." These lectures were summarized by Morris Cohen, a professor of philosophy at CCNY.[14] In his introductory remarks, Professor Cohen characterized Einstein's presentation, saying:

> Particularly illuminating and suggestive was the way Prof. Einstein showed that great and fundamental achievements in science are not brought about by tinkering with details and effecting petty improvements, but rather by radical re-examination of fundamental ideas and trusting the scientific imagination in the suggestion of new departures or lines of development.[15]

The summary of the third lecture quotes Einstein's conclusion:

> The general significance of the theory of Relativity he held to consist in the fact that it shows us the folly of beginning with Euclidean space and then asking how gravity acts in such a world. Instead, we should begin with the study of the way bodies actually behave and formulate the nature of space accordingly.[16]

Einstein's next academic stopover was in Chicago, where, during the first week of May, he delivered three lectures at the University of Chicago and one at a local school, the Francis W. Parker School. Einstein hoped that his visit would give him the opportunity to meet local scientists, specifically Robert A. Millikan and Albert A. Michelson. Michelson was away, traveling in Europe, but Millikan, who had not yet taken up his position as president of the California Institute of Technology (Caltech), was present. What remains of the three lectures delivered at the university is a typescript of shorthand notes, in German, prepared by a stenographer.[17] These notes are roughly similar, though not identical, to the available notes of the two popular lectures that Einstein would later give in Princeton, which are quite different from his professional talks in Princeton. The brief record of Einstein's lecture at the Parker school indicates that he may have discussed there his route to the general theory of relativity at greater length than the single sentence available: "Then I made the second discovery of my life, that is, the discovery of the equivalence of inertia and of gravity."[18]

Einstein's first lecture in Princeton on 9 May was preceded by a ceremony conferring upon Einstein the honorary degree of doctor of science. President Hibben said in his opening remarks:

> In your writing you have in your daring thought speculated upon the possibility of a finite and yet unlimited universe. Whether the world is finite or infinite is not for me to say. Certainly, however, there is a world, which has no limits whatever. This is the world of the spirit, to which you belong by unquestioned right.

In his presentation address, the dean of the university, Andrew West, referred to Einstein as a legitimate successor of giants of science like Maxwell, Newton, Galileo, and Pythagoras and as "the foremost scientific theorist of our age." He was more specific, saying:

FIGURE 2: The honorary degree procession from the ceremony to the lecture hall at Princeton University.

> But his highest intellectual fame rests on his general theory of relativity, a new, profound and daring conception of space and time and of their relation to the physical universe, affecting our theories of light and gravitation and joining the two mysteries of inertia and gravitation in one, yet more elemental conception.

West concluded his remarks by mentioning Einstein's courageous public expression of opinion at the beginning of WWI in Germany: "And especially would we do him honor for the moral fidelity whereby, amid distressing perils, he refused to join with others in condoning the invasion of Belgium."[19]

This last remark refers to the "To the Civilized World," often referred to as the Manifesto of the 93, which conveyed the enthusiasm of broad intellectual circles in Germany.[20] This document expressed loyalty to the German forces and rejected accusations of brutal behavior and of deliberate destruction of treasures of art and science. It asserted that "were it not for German militarism, German culture would have been wiped off the face of the earth."[21] Einstein refused to sign this appeal. Together with the physiologist Georg Nicolai, he drafted the "Manifesto to the Europeans," which constituted a pacifist response to the Manifesto of the 93. It implicitly criticized the nationalist passions conveyed by that document and appealed to artists and scientists not to abandon the "common world civilization."[22] The memories of WWI were still fresh, and the dean found it appropriate to highlight this episode in his honorary-degree presentation speech.

The local newspapers reported in great detail on every aspect of Einstein's visit, often under sensationalistic headlines. They printed summaries of every lecture and highlighted specific statements and predictions by Einstein, such as his admission that the theory

would collapse if the spectral lines of the light coming from the sun were not shifted toward the red side of the spectrum (the gravitational redshift): "If this is untrue the whole theory is a fallacy." The journalists reported on Einstein's fascination with student life in the dormitories and about his impression of great progress in scientific research in the United States compared with the postwar stagnation in Europe. They interviewed scholars in different fields about their views on Einstein's theory. Here are two examples.

The New York Evening Post, on May 13, interviewed Prof. E. G. Spaulding from the department of philosophy at Princeton University on his views on Einstein's theory: "*The Einstein theories have no bearing at all upon problems of ethics, problems of psychology, problems of theology, or, in general, upon the most important ones in philosophy*." This new theory therefore does not upset the laws of morality, and the principle of relativity, as Prof. Spaulding points out, does not apply to evaluation of goodness.

Prof. Spaulding regards the Einstein principles as significant, not only because of their importance in the spheres of physics and mathematics, but because they demonstrate that science as a whole, which has prided itself upon dealing with facts in an unbiased manner . . . is not free from the fault of proceeding upon preconceived notions without sufficient experimental verification.[23]

The New York Evening Post, on May 12, reported on the response of Prof. L. P. Eisenhart from the department of mathematics at Princeton University to the query whether Einstein's speculations (such as that on the rotating disk) are "*much ado about nothing*." The professor said: "*The question has been raised whether the Einstein theory is any good at all. . . . As further experimental results are obtained it may be necessary to change the physical interpretation, which the genius of Einstein gave to a mathematical structure by a group of brilliant geometers from the time of Riemann. But Einstein's conception of the relation between geometry and physics is likely to exert a lasting effect upon theories concerning the physical world*."[24]

NOTES

1. Luther P. Eisenhart to Einstein, 1 October 1920, CPAE vol. 10, Doc. 160.
2. Einstein to John G. Hibben, 14 November 1920, CPAE vol. 10, Doc. 203.
3. Paul Ehrenfest to Einstein, 7 November 1920, CPAE vol. 10, Doc. 191.
4. John G. Hibben to Einstein, 24 December 1920, CPAE vol. 10, Doc. 243.
5. Einstein to John G. Hibben, 21 February 1921, CPAE vol. 12, Doc. 53.
6. Einstein to Hendrik A. Lorentz, 22 February 1921, CPAE vol. 12, Doc. 57.
7. Hendrik A. Lorentz to Einstein, 19 March 1921, CPAE vol. 12, Doc. 107.
8. Fritz Haber to Einstein, 9 March 1921, CPAE vol. 12, Doc. 87.
9. Einstein to Fritz Haber, 9 March 1921, CPAE vol. 12, Doc. 88.
10. Vol. 7, 56b, "Professor Einstein on the Proposed Hebrew University of Jerusalem," CPAE, vol. 13, p. 45.
11. Quotation from József Illy, *Albert Meets America: How Journalists Treated Genius during Einstein's 1921 Travels*, annotated ed. (Baltimore, MD: John Hopkins University Press, 2006), p. 102.

12. Chaim Weizmann to Einstein, 9 March 1921, CPAE, vol. 12, Doc. 91.
13. See Illy, *Albert Meets America*.
14. *The Campus*, 29 April 1921 (CCNY), reprinted as Appendix C, CPAE, vol. 12 (German ed.), pp. 507–512.
15. *The Campus*, p. 507.
16. Ibid., p. 510.
17. Reprinted as Appendix D, CPAE, vol. 12 (German ed.), pp. 513–528.
18. Appendix D, CPAE, vol. 12 (German ed.), p. 519. Translation by H. G.
19. Quotations from *Princeton Alumni Weekly*, 11 May 1921, pp. 713–714.
20. "An die Kulturwelt! Ein Aufruf," http://germanhistorydocs.ghi-dc.org/pdf/deu/817_Bernhard_vom_Brocke_156.pdf. In English, http://www.jstor.org/stable/pdf/25122278.pdf?acceptTC=true.
21. Quoted in Walter Isaacson, *Einstein: His Life and Universe* (New York: Simon & Schuster, 2007), p. 207.
22. CPAE, vol. 6, Doc. 8, p. 28.
23. Illy, *Albert Meets America*, p. 196.
24. Ibid., p. 189.

3
STRUCTURE AND CONTENTS OF *THE MEANING OF RELATIVITY*

It has already been mentioned that this book is not a systematic page-by-page, or even chapter-by-chapter, commentary on Einstein's text. Thus, it may be useful to provide at the outset a brief summary of *The Meaning of Relativity* and of the appendixes added to later editions. This chapter is intended mainly to show the flow of ideas in the order of their presentation and to emphasize the new ways of their formulation, as influenced by Einstein's interaction with his colleagues and by his own rethinking of some of the basic concepts. Many of the points and topics mentioned in this chapter will be recapitulated in part III, in chapters devoted to specific topics.

As in the case of his *Relativity—The Special and the General Theory*, where throughout the years Einstein added new appendixes to later editions of the book, he also used *The Meaning of Relativity* as a platform to publish over the course of time, up until the end of his life, his reflections on the major issues on the agenda of debates and efforts during the formative years of general relativity. We have already mentioned in the introduction that the appendixes were added to the different editions between 1945 and 1956. These appendixes will be discussed in the relevant chapters in part III; therefore, they are not summarized here.

The Meaning of Relativity is based on the lectures delivered in Princeton—specifically on the three professional lectures: "Foundations of the Special Relativity Theory," "Foundations of the General Relativity Theory," and "The Cosmological Problem." We do not have scripts of these lectures as we do for the first two, popular, lectures, entitled "Generalities on the Theory of Relativity," which were attended by a nonspecialist audience. All we know about them is from the brief summaries that were written by Edwin P. Adams and published in the popular press and from unofficial abstracts of these lectures, which also contain a summary of the mathematical formalism.[1] Comparing these summaries with the text of Einstein's book, we occasionally find quite significant differences in scope and presentation, some of which we shall point out in our commentaries.

Space and Time in Pre-relativity Physics. The first chapter of the *Meaning of Relativity* was not delivered as a lecture in Princeton in May 1921. It was written later, sometime in September, specifically for the published version.[2] It begins with an epistemological

discussion of space and time and of geometry in relation to Einstein's formulation of relativity theory. (These issues are explored in detail in chapter 1 of part III.) Following this discussion, Einstein introduces the concept of invariants. These are the only quantities to have an objective significance with respect to a specific geometry. In Euclidean geometry, which is the hallmark of pre-relativity physics, such quantities are independent of a particular choice of Cartesian coordinates. Two examples of such invariants are the length of a line segment and the volume element. Einstein then introduces the basics of tensor calculus. It is the first time that he does this in the context of classical physics. Tensor formulation of the laws of physics guarantees that they are covariant with respect to transformations, which represent the symmetry of space. The symmetry of space, namely, its isotropy and homogeneity, determines the form of physical laws. Einstein refers to this as "the principle of relativity with respect to direction" (p. 184 [24]). He then demonstrates how this principle is reflected in some of the equations of physics. As examples he uses the equation of motion of a particle in a force field, the equations of motion of a continuous medium, and Maxwell's equations.

The discussion of the tensor formulation of the laws of physics in this chapter is an enlightening prelude to the role of these concepts in special, and even more so, in general relativity.

The Theory of Special Relativity. The transition to the theory of special relativity is marked by the extension of "the principle of relativity with respect to direction" (p. 184 [24]) to include also "relativity with respect to translation," namely, with respect to the state of motion of the space of reference. In classical mechanics, reference frames moving at uniform velocity with respect to each other (inertial frames) are physically equivalent. The transformation between the coordinates in two inertial reference frames is based on the assumption that the time between two events and the length of a segment are absolute. A different transformation is needed to guarantee the equivalence of inertial frames not only to mechanical phenomena but to all physical phenomena (specifically to electromagnetic phenomena). This goal is called the *principle of special relativity*. Einstein then shows how this principle of relativity and the principle of the constancy of the velocity of light uniquely determine this transformation: the Lorentz transformation. Unlike in pre-relativity physics, space and time are not separate entities. They are combined into a four-dimensional spacetime continuum, which is most clearly demonstrated in the formulation of the theory of special relativity due to Hermann Minkowski. However, one should keep in mind that the time coordinate is physically different from the space coordinates; they appear with opposite signs in the expression of the distance between two points (two events) in the spacetime continuum. Einstein then defines vectors and tensors with respect to the Lorentz transformation, leading to a tensor formulation of Maxwell's equations, of the electromagnetic force acting on charges and currents, and of the energy-momentum of the electromagnetic field and its conservation law. He also introduces the energy-momentum tensor of matter and discusses the basic equations of the hydrodynamics of perfect fluids (the Euler equations).

The General Theory of Relativity. The first part of this chapter is devoted to a discussion of the transition from the special to the general relativity principle. According to Einstein's understanding of such a general relativity principle, physical laws are independent of the state of motion of the reference space in which they are described.

Einstein argued that such a generalization of the relativity principle to include accelerated reference frames is possible because all inertial effects caused by acceleration can be alternatively attributed to the presence of a gravitational field. This is the essence of the equivalence principle, which Einstein had already formulated in 1907. Thus, a general theory of relativity is at the same time a relativistic theory of gravitation. As he had done several times before, Einstein then uses the model of a rotating disk to show that general relativity implies non-Euclidean geometry and that the gravitational field is represented by curved spacetime. This is expressed by the dual role of the metric tensor $g_{\mu\nu}$. Its components describe the geometry of spacetime and at the same time determine the gravitational potential.

After the introduction of these basic concepts and principles, Einstein is ready to present the mathematical formulation of the theory. What is presented here is different from the mathematical formulation in the seminal 1916 paper. About two months after his final presentation of the theory to the Royal Prussian Academy of Science, in November 1915, Einstein wrote to Lorentz:

> My series of gravitation papers are a chain of wrong tracks, which nevertheless did gradually lead closer to the objectives. That is why now finally the basic formulas are good, but the derivations abominable; this deficiency must still be eliminated.[3]

Without eliminating what appeared to him to be an avoidable complexity, in March 1916 he submitted the manuscript for publication to *Annalen der Physik*, the leading physics journal at that time. The "abominable derivations" and the use of special coordinates, which marked Einstein's earlier presentations of his general theory, have been elegantly corrected in this and in the following chapter of the *Meaning of Relativity.*

The exposition of tensor calculus and the discussion of the metric tensor here is simpler and more transparent. But what is more important is Einstein's treatment of differential geometry in its physical context. The mathematical framework of his general theory of relativity emerged from the absolute differential calculus of Elwin Christoffel, Gregorio Ricci-Curbastro, and Tullio Levi-Civita. But in his original formulation of general relativity, Einstein did not systematically introduce non-Euclidean geometry, nor did he interpret his own theory in terms of differential geometry. When he discussed the Riemann-Christoffel tensor, for instance, he did not even mention curvature.

The relation of Einstein's theory to differential geometry and its geometrical interpretation was developed only after the establishment of Einstein's theory and was not a presupposition in its formulation. The geometrization of general relativity and the understanding of gravity as being due to the curvature of spacetime is a result of the further development. It was only in the Princeton lectures that Einstein acknowledged covariant differential operations on tensors as being most satisfactorily performed by the method introduced by Levi-Civita, which was based on the notions of *parallel transport* and *affine connection* (see box), and later used in the context of general relativity by the mathematician and philosopher Hermann Weyl, as acknowledged by Einstein (p. 229 [69]). It was Weyl, in particular, who clarified the geometrical interpretation of the Riemann-Christoffel curvature tensor and related it to the parallel displacement of a vector around a closed loop.

THE AFFINE CONNECTION

Mathematicians in the nineteenth century began to explore the notion of connection, which describes how geometrical data are consistently transported along specific curves. The most elementary type of connection and the most relevant for our discussion is the affine connection, which specifies how vectors are transported in parallel, along a curve, from one point to another. The affine connection is intimately related to the derivative of a vector in a certain direction, namely, to the question of how a vector is changed by an infinitesimal transport in a given direction from point P to a neighboring point P′. In Riemannian geometry, only the direction of the vector changes; its length remains constant. The difference between the shifted vector and the original vector may be regarded as the differential of the vector at P. Although Einstein does not use the term "affine connection," this is exactly what it is (p. 70, eq. 67). Levi-Civita used the concept of parallel transport to clarify the concept of covariant differentiation. In this context, the affine connection is also known as the Levi-Civita connection. Its mathematical expression is given by the Christoffel symbol Γ—a combination of first derivatives of components of the metric tensor.

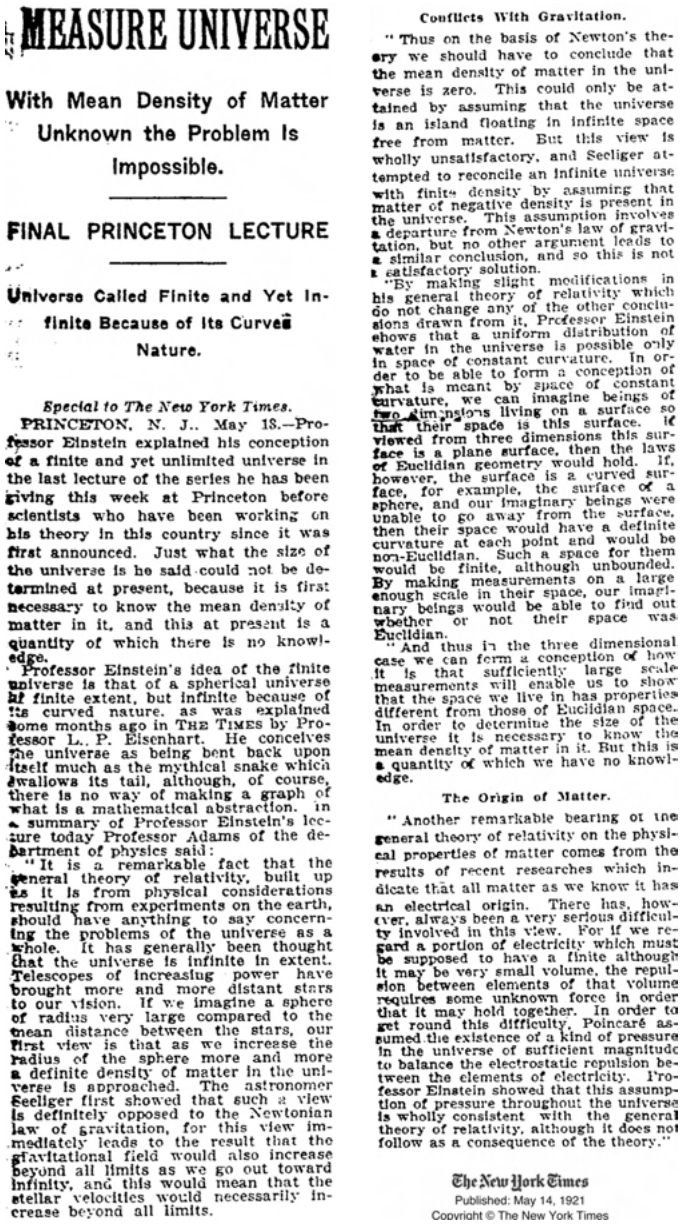
Einstein learned about these concepts and this method, which he used, from Weyl's book *Space, Time, Matter*. [4] He introduced them in his lectures on general relativity in Berlin 1919.[5] He applied this method fully for the first time in a published text in *The Meaning of Relativity*. Weyl instructed his publisher to send Einstein printer's proofs of his book. Einstein's reaction was very enthusiastic:

> I am reading with genuine delight the correction proofs of your book. . . . It is like a symphonic masterpiece. Every word has its relation to the whole and design of the work is grand. What a magnificent method the infinitesimal parallel displacement of vectors is for the deriving of the Riemann tensor! How naturally it all comes out.[6]

The notion of parallel transport of a vector can also be applied to the geodesic line. The equation of this line was derived in Einstein's original formulation as the line of extremal length between two points. It can also be defined, as is done here, as the line for which the tangent vector remains parallel to itself when moved from point to point along the line. The tangent vector is a unit vector along the tangent direction at a given point on the line. Intuitively, this requirement means that the geodesic line is the "straightest" possible line between two points.

In this chapter, Einstein has developed and described the mathematical apparatus that is necessary to formulate the physical contents of the general theory of gravity, which is done in the next chapter.

The General Theory of Relativity (Continued). The equation of the geodesic line, derived at the end of the previous chapter, also defines the trajectory of motion of a material particle under the action of gravitation (without any other forces). Einstein then shows that in the limit of weak fields and low velocities, this equation reduces to Newton's equation of motion. He proceeds to derive the gravitational field equation. The guiding model is Poisson's equation of Newtonian physics, which shows how the gravitational field (the



MEASURE UNIVERSE

With Mean Density of Matter Unknown the Problem Is Impossible.

FINAL PRINCETON LECTURE

Universe Called Finite and Yet Infinite Because of Its Curved Nature.

Special to The New York Times.

PRINCETON, N. J., May 13.—Professor Einstein explained his conception of a finite and yet unlimited universe in the last lecture of the series he has been giving this week at Princeton before scientists who have been working on his theory in this country since it was first announced. Just what the size of the universe is he said could not be determined at present, because it is first necessary to know the mean density of matter in it, and this at present is a quantity of which there is no knowledge.

Professor Einstein's idea of the finite universe is that of a spherical universe of finite extent, but infinite because of its curved nature, as was explained some months ago in THE TIMES by Professor L. P. Eisenhart. He conceives the universe as being bent back upon itself much as the mythical snake which swallows its tail, although, of course, there is no way of making a graph of what is a mathematical abstraction. In a summary of Professor Einstein's lecture today Professor Adams of the department of physics said:

"It is a remarkable fact that the general theory of relativity, built up as it is from physical considerations resulting from experiments on the earth, should have anything to say concerning the problems of the universe as a whole. It has generally been thought that the universe is infinite in extent. Telescopes of increasing power have brought more and more distant stars to our vision. If we imagine a sphere of radius very large compared to the mean distance between the stars, our first view is that as we increase the radius of the sphere more and more a definite density of matter in the universe is approached. The astronomer Seeliger first showed that such a view is definitely opposed to the Newtonian law of gravitation, for this view immediately leads to the result that the gravitational field would also increase beyond all limits as we go out toward infinity, and this would mean that the stellar velocities would necessarily increase beyond all limits.

Conflicts With Gravitation.

"Thus on the basis of Newton's theory we should have to conclude that the mean density of matter in the universe is zero. This could only be attained by assuming that the universe is an island floating in infinite space free from matter. But this view is wholly unsatisfactory, and Seeliger attempted to reconcile an infinite universe with finite density by assuming that matter of negative density is present in the universe. This assumption involves a departure from Newton's law of gravitation, but no other argument leads to a similar conclusion, and so this is not a satisfactory solution.

"By making slight modifications in his general theory of relativity which do not change any of the other conclusions drawn from it, Professor Einstein shows that a uniform distribution of water in the universe is possible only in space of constant curvature. In order to be able to form a conception of what is meant by space of constant curvature, we can imagine beings of two dimensions living on a surface so that their space is this surface. If viewed from three dimensions this surface is a plane surface, then the laws of Euclidian geometry would hold. If, however, the surface is a curved surface, for example, the surface of a sphere, and our imaginary beings were unable to go away from the surface, then their space would have a definite curvature at each point and would be non-Euclidian. Such a space for them would be finite, although unbounded. By making measurements on a large enough scale in their space, our imaginary beings would be able to find out whether or not their space was Euclidian.

"And thus in the three dimensional case we can form a conception of how it is that sufficiently large scale measurements will enable us to show that the space we live in has properties different from those of Euclidian space. In order to determine the size of the universe it is necessary to know the mean density of matter in it. But this is a quantity of which we have no knowledge.

The Origin of Matter.

"Another remarkable bearing of the general theory of relativity on the physical properties of matter comes from the results of recent researches which indicate that all matter as we know it has an electrical origin. There has, however, always been a very serious difficulty involved in this view. For if we regard a portion of electricity which must be supposed to have a finite although it may be very small volume, the repulsion between elements of that volume requires some unknown force in order that it may hold together. In order to get round this difficulty, Poincaré assumed the existence of a kind of pressure in the universe of sufficient magnitude to balance the electrostatic repulsion between the elements of electricity. Professor Einstein showed that this assumption of pressure throughout the universe is wholly consistent with the general theory of relativity, although it does not follow as a consequence of the theory."

The New York Times
Published: May 14, 1921
Copyright © The New York Times

FIGURE 3: Summary by Professor Adams of the final Princeton lecture. *New York Times*, 14 May 1921.

left-hand side of the equation) is generated by the density of matter (the right-hand side). In general relativity, the scalar density of matter is replaced by a tensor representing the energy and momentum of matter. This tensor generates the left-hand side of the equation, which represents the geometry of spacetime corresponding to the gravitational field. It is determined by the requirement of energy-momentum conservation, which leads to equation 96. Applying this equation to the case of weak fields and low velocities, Einstein

shows that in the first approximation the gravitational field differs fundamentally from Newton's theory as represented by Poisson's equation. This is because the gravitational potential is a tensor and not a scalar. Still, the physical behavior corresponds to the Newtonian limit because only one component of that tensor appears in the equation of motion of a material particle.

Next, Einstein uses the gravitational field equation to derive the three effects that served as the first tests of the theory: the bending of light by the gravitational field of the sun, the shift to the red of spectral lines emitted by atoms in a gravitational field (gravitational redshift), and the motion of the perihelion of planet Mercury. Judging from the available summaries of the last Princeton lecture, this part of the chapter was apparently written specifically for the printed version, although these effects are discussed without the mathematical formalism in the popular Princeton lectures. The derivation of the first two effects is the same as in Einstein's original version of the theory. The third effect was originally calculated by a method of successive approximations, which Einstein had developed previously. He could have derived it using the exact solution for a radially symmetric and static gravitational field found by Karl Schwarzschild. Einstein was aware of this solution at the time of writing his seminal article "The Foundation of General Relativity." Here he uses the Schwarzschild solution and refers to an elegant derivation by Hermann Weyl.

After a short remark about expressing Maxwell's equations of the electromagnetic field, Einstein turns to the second part of this chapter, which is devoted to the "so-called cosmological problem" (p. 258 [98]). Its main theme is the defense of Mach's principle, which states that all inertial phenomena, namely the fictitious forces arising in accelerated reference frames, are caused by all the masses in the universe. This idea was Einstein's heuristic guideline in his search for a theory of relativity. In an ongoing dialogue on this idea and related issues, it was criticized by his colleague and discussion partner, the Dutch astronomer Willem de Sitter. Although de Sitter's name is not mentioned, it is clear that the text is motivated by the debate between the two of them. Einstein's discussion leads, as a conclusion, to his model of a spherical, static, and closed universe. This part of the chapter will be discussed in detail in chapter 4 of part III.

NOTES

1. CPAE, vol. 12, Appendix E, pp. 529–538.
2. The editors of CPAE 7 present two texts (Docs. 63 and 64) that are dated on the assumption that they are drafts of Einstein's *The Meaning of Relativity*.
3. Einstein to Hendrik A. Lorentz, 17 January 1916, CPAE vol. 8, Doc. 183.
4. Hermann Weyl, *Space, Time, Matter* (New York: Dover Publ., 1950), 1. Reprint of the 4th ed. [1922].
5. CPAE vol. 7, Doc. 19.
6. Einstein to Hermann Weyl, 8 March 1918, CPAE vol. 8, Doc. 476.

PART II

THE EMERGING WORLD OF GENERAL RELATIVITY

1
PHYSICS AND GEOMETRY

The first chapter of The Meaning of Relativity *begins with epistemological* remarks about space and time. Einstein developed his views on geometry in close relation with his formulation of relativity theory. Indeed, the first sentence of the book reads: "The theory of relativity is intimately connected with the theory of space and time." It is common knowledge that the two theories of relativity have fundamentally changed our notion of space and time. In January 1921, Einstein delivered a lecture to the Prussian Academy of Sciences entitled "Geometry and Experience," which has become the canonical text reflecting the main points of his thinking on this subject.[1] He discussed the concepts of space and time on several occasions before that and continued to do so almost until the end of his life. An extensive analysis of these concepts is found in his *Relativity—The Special and General Theory*.[2] To the last edition of that book, which appeared a year before his death, Einstein added a long appendix; it summarizes the development of his perception of space and may be viewed as his epistemological legacy on this issue.

A. PRACTICAL GEOMETRY

Einstein distinguished between *axiomatic geometry*, a purely mathematical theory comprising propositions logically derived from a set of axioms, and *practical geometry*, which is related to experience. According to him, mathematics in general, and particularly geometry, has been embraced by mankind because of the need to learn something about the behavior of real objects. Pure axiomatic geometry alone cannot help to make assertions about the behavior of real objects, however, but is necessary for the process of surveying. The German mathematician David Hilbert set up an axiomatic foundation of Euclidean geometry based on three systems of "things," namely, points, straight lines, and planes. He made no assumption about the nature of these "things." He allegedly said that one can replace points, straight lines, and planes at any time with tables, chairs, and beer mugs, as long as the axioms are satisfied. What Einstein calls practical geometry comprises, in contrast, such a relation to real objects. He explicitly adds the concept of rigid bodies to the conceptual edifice of geometry, thus turning it into a physical science in which truth can be established by agreement with reality.

These ideas are expressed in detail and with a decisive message about their significance for the theory of relativity in Einstein's "Geometry and Experience." There he concludes: "Geometry thus completed is evidently a natural science; we may in fact regard it as the most ancient branch of physics." He emphasizes the importance of this statement: "I attach special importance to the view of geometry which I have just set forth, because without it

I should have been unable to formulate the theory of relativity."[3] This statement is related to Einstein's taking issue with Poincaré's conventionalist philosophy of science. We shall substantiate this message below in section D, basing our arguments on the painstaking analysis by the philosopher of science Yemima Ben-Menachem.

Until 1921, Einstein did not mention Poincaré explicitly. We know that he read Poincaré's *Science and Hypothesis* with his friends in the Akademie Olympia in 1902. Einstein's biographer Abraham Pais quotes one of the members, Maurice Solovine, as saying: "This book profoundly impressed us and kept us breathless for weeks on end."[4] In *The Meaning of Relativity*, Einstein refers to Poincaré several times. This may have been due to his increased interest in the philosophical aspects of the theory of relativity. At the beginning of the discussion on space (p. 163 [3]), he praises Poincaré's insight that our perception of space is developed through the operations we can perform on rigid bodies and quotes his statement that "among all the changes we can perceive in a rigid body those which can be cancelled by a voluntary motion of our body are marked by their simplicity" (p. 163 [3]). This is the beginning of a discussion of concepts such as "bodies of reference" or "spaces of reference," Cartesian coordinates associated with spaces of reference and Euclidean geometry, and it already alludes to symmetry with respect to orientation in space.

Einstein begins his discussion of geometrical concepts by referring to the philosopher Immanuel Kant, but without mentioning his name. Kant argued that certain concepts and assertions are part of our a priori knowledge, that is, independent of and prior to any experience. More specifically, geometry was for him an example of synthetic a priori truths and therefore both necessary and automatically applicable to experience. More specifically, Euclidean geometry forms our intuitive perception of spatial relations between material bodies without which we could not even imagine them. Einstein categorically rejects this position:

> I am convinced that the philosophers have had a harmful effect upon the progress of scientific thinking in removing certain fundamental concepts from the domain of empiricism, where they are under our control, to the intangible heights of the *a priori*. . . . This is particularly true of our concepts of time and space, which physicists are obliged by the facts to bring down from the Olympus of the *a priori* in order to adjust them and put them in a serviceable condition. (p. 162 [2])

Although it is clear that this passage refers to Immanuel Kant's notion of "a priori,"[5] with his statement about the harmful effect of philosophers on scientific thinking, Einstein might also have had in mind Henri Bergson's concept of time. Half a year after summarizing his Princeton lectures, both of them engaged in a public debate in which Einstein bluntly stated that Bergson's philosophical time does not exist.[6]

Einstein leads the reader through a discussion of known elementary concepts in order to demonstrate that some of the more sophisticated ideas of relativistic physics are already present in pre-relativistic physics:

- The location of a point marked on a rigid body (the *body of reference*) is characterized by three numbers (*coordinates*).
- Two points on such a rigid body form an interval. The length of an interval may be measured using standard measuring rods.
- The simplest mathematical expression of the distance between two points is given by Cartesian coordinates (eq. 1).

- The bodies of reference can be moved and oriented arbitrarily in space.
- If the distance between two points on a body of reference does not depend on its position, the space is Euclidean.
- The invariance of the distance implies that the Cartesian coordinates in two positions in space are related to each other by an orthogonal transformation (eq. 3).

To summarize the essence of the discussion up to this point, let us use Einstein's own words:

> In the Euclidean geometry there are (in a given space of reference) preferred systems of coordinates, the Cartesian coordinates, which transform into each other by linear orthogonal transformations. The distances between two points of our space of reference, measured by a measuring rod, is expressed in such coordinates in a particularly simple manner. . . . In the present treatment, geometry is related to actual things (rigid bodies), and its theorems are statements concerning the behavior of these things, which may prove to be true or false. (p. 167 [7])

In concluding this discussion, Einstein again emphasized the distinction between the mathematical and the physical conceptions of geometry:

> The questions as to whether Euclidean geometry is true or not does not concern him (the mathematician). But for our purpose it is necessary to associate the fundamental concepts of geometry with real objects: without such association geometry is worthless for the physicist. The physicist is concerned with the question as to whether the theorems of geometry are true or false. (p. 168 [8])

B. THE "PRINCIPLE OF RELATIVITY WITH RESPECT TO DIRECTION"

To ensure that the laws of physics are not affected by changes of position in space of the reference system, "It is postulated that the system of equations expressing the laws of pre-relativity physics is co-variant with respect to the transformation (eq. 3), as are the relations of Euclidean geometry" (p. 176 [16]).

To guarantee the covariance of the laws of physics, they should be expressed by tensors. Tensors are rather complicated mathematical objects that are indispensable in general relativity. In preparation for the forthcoming lectures, Einstein already introduces tensor analysis and the concept of "covariance" in the chapter on pre-relativity physics. What is the significance and the role of tensors?

The laws of physics are formulated as mathematical equations. The physical entities on both sides of these equations are represented by mathematical expressions, which are functions of position in space and time, and they assume different values in different reference frames, as represented by different coordinate systems. The requirement of covariance means that even though the two sides of an equation may change upon transition from one reference frame to another, their equality is maintained for all *bodies of reference*, irrespective of their orientation in space. For orthogonal transformations, such as ordinary rotations in space, the mathematical entities that have this property are tensors. Vectors and scalars are special cases of tensors (tensors of rank 1 and 0). In pre-relativity physics, tensors (of rank 2) are represented by 3×3 matrixes of functions of space coordinates.

Einstein refers to the covariance in pre-relativity physics as the "principle of relativity with respect to direction" (p. 184 [24]). It is important to emphasize the role of symmetry. The symmetry of Euclidean space, including the invariance to rotations in space, constrains the admissible laws of nature. This will become even more apparent in section C.

C. THE FOUR-DIMENSIONAL SPACETIME CONTINUUM

The bodies of reference of the previous sections are now allowed to move at constant velocities with respect to each other, thus forming inertial frames of reference. The physical equivalence of different inertial frames of reference is the essence of the special theory of relativity and composes the content of the second Princeton lecture.

The consequences of this theory could be conveniently described in the framework of a new mathematical formalism developed by Hermann Minkowski, who was a professor of mathematics at the Federal Technical University in Zurich when Einstein was a student there and attended several of Minkowski's courses. In 1908, he showed that Einstein's special theory of relativity could be understood geometrically as a theory of four-dimensional spacetime.

This formalism combines space and time into one entity—spacetime—and assigns a geometrical distance between any two physical events occurring at different positions and different times. One usually refers to points in spacetime as events because they are characterized by the location and time of occurrence. Minkowski's spacetime is equipped with a "metric" instruction, which is employed to measure the distance between two events and also characterizes their causal relation in the sense of whether one event can causally affect another. The square of this distance is simply the square of their spatial separation minus the square of the time separation (multiplied by c^2) between the two events. Properly speaking, it is not an ordinary metric. The minus sign preceding the time component makes it a metric of a pseudo-Riemannian manifold, which is a manifold in which the distance between two distinct points may be zero. Observers moving at constant velocity with respect to each other may compute this number using their respective rods and clocks, and they will obtain the same result.

It took Einstein some time to appreciate Minkowski's geometric formulation of the theory of special relativity as an interesting and useful contribution. Einstein used Minkowski's formulation in the context of special relativity for the first time in his unpublished review of 1912.[7] By that time, he had become aware of the significance of von Laue's work on relativistic continuum theory, which elaborated the dynamical implications of special relativity and, in particular, the understanding of energy-momentum laws when cast into a four-dimensional formalism.[8] Minkowski's formulation of the theory became the framework of its later development and led Einstein to his theory of general relativity. In the first paragraph of Einstein's seminal article "The Foundation of General Relativity," published in March 1916, he wrote: "The generalization of the theory of relativity has been facilitated considerably by Minkowski, a mathematician who was the first one to recognize the formal equivalence of space coordinates and the time coordinate, and utilized this in the construction of the theory."[9]

The condition of invariance of the distance between two events implies that the spacetime coordinates in two reference frames are related by the *Lorentz transformation*. It was derived by Hendrik Lorentz as the transformation that leaves the Maxwell equations the same in all inertial frames of reference. Einstein derived this transformation in *The Meaning*

of Relativity as the transformation that leaves the spacetime interval invariant. If one thinks about his enduring legacy, which transcends the specifics of the theory he formulated at that time, his major contribution was to revolutionize how we view symmetry. It is Einstein who prepared the ground for our present-day perspective of symmetry. He put symmetry first and regarded it as the primary feature of nature, because it constrains the allowed physical laws. Thus, the transformation properties of the electromagnetic field were not to be derived from Maxwell's equations, as Lorentz had done, but are consequences of the spacetime distance invariance, a property of the Minkowski spacetime continuum, and they essentially dictate the form of Maxwell's equations and all other laws of physics. This is a profound change of attitude, one that has been followed with great success in physics since that time.

Moreover, Einstein's perception of the role of symmetry in general relativity was a source of inspiration for the emergence of Emmy Noether's theorems, which deepened our general understanding of the role of symmetry in physics. As early as 1914, Einstein, in the context of his preliminary *Entwurf* theory,[10] had found that the same conditions that determine the covariance of the Lagrangian also ensure energy-momentum conservation (see box).[11] His result is a special case of a theorem that was published in 1918 by Emmy Noether and connects symmetries and conservation laws.[12] Emmy Noether was a German mathematician working alongside David Hilbert at the University of Göttingen, one of the great centers of mathematics at that time. Noether is known for her groundbreaking contributions to abstract algebra and theoretical physics. In physics, she is best known for what are now called Noether's theorems. According to them, every symmetry in nature is associated with a corresponding conservation law, and they establish a deep connection between the underlying geometry of the universe and the behavior of energy and matter. The mathematical novelty of these theorems was that she showed how Lie's theory of differential equations can be applied to the Euler-Lagrange equations derivable from an invariant action integral. The basic idea was due to the mathematician Felix Klein, and Emmy Noether showed how it could be made to work for both coordinate and field variables.

THE LAGRANGIAN FORMALISM

Newton's mechanics is based on the concept of force, which is mathematically represented by a vector. Beginning with the work of Leibniz and its extension by Euler, Lagrange, and Hamilton, an alternative approach emerged that extends well beyond mechanics. It is based on the characterization of a physical process, such as the motion of a particle, by a single quantity, usually called *Lagrangian*, which depends on the parameters describing the state of the system and their derivatives with respect to the space (or spacetime) coordinates. Instead of an equation of motion, the dynamics is described by a variational principle. According to the procedure, which was first introduced by Hamilton, the initial and final points of a motion are fixed and the possible paths between them are characterized by a certain quantity, referred to as the action, which may be obtained by the time integral of the Lagrangian. The actual motion of the particle (or the dynamics of the physical system) is given by an extremal (minimum or maximum) value of this integral. From Hamilton's *variational principle*, it is then possible to derive certain differential equations for the motion, the so-called *Euler-Lagrange equations*. The Lagrangian formalism is one of the main tools for the description of the dynamics of a vast variety of physical systems.

Noether's theorems have been considered as one of the most important mathematical contributions guiding the development of modern physics. In the context of general relativity, general covariance can be interpreted as a basic symmetry characterizing the spacetime geometry of the universe, and energy-momentum conservation is the associated conservation law implied by Noether's theorems. After Noether's death in 1935, Einstein wrote about her in a letter to the *New York Times*: "In the judgment of the most competent living mathematicians, Fräulein Noether was the most significant creative mathematical genius thus far produced since the higher education of women began."[13]

D. THE NON-EUCLIDEAN CONTINUUM

The transition from special to general relativity involved a generalization from a (pseudo) Euclidean to a non-Euclidean continuum. The main point in this transition is the role of coordinates. In Euclidean geometry one can assign a physical meaning to the Cartesian coordinates: they label the corners of a grid of identical measuring rods, which covers a plane or the three-dimensional space. Thus, they are directly connected to distance measurements. This physical meaning of the coordinates has to be abandoned in the discussion of the non-Euclidean continuum of general relativity because the relation between the coordinates and physically meaningful distances is mediated by a complex mathematical object, the metric tensor, which itself is a function of the coordinates. Einstein discussed this point in detail in his book *Relativity—The Special and General Theory.*"[14]

In that "booklet," Einstein began the discussion of the non-Euclidean continuum by exploring a famous example that goes back Poincaré, although Einstein does not mention him. In this example, a flat surface, such as the surface of a marble table, is covered by a square grid of measuring rods of equal length. This structure forms a two-dimensional Euclidean continuum with the corners of the grid as Cartesian coordinates. Let the table be heated in the middle so that the measuring rods in the warmer areas of the surface expand. It turns out that with the help of these measuring rods, one can no longer create the familiar coordinate grid that can normally be imposed on such a flat surface to label its points. The observer innocently using these rods and confronted with this problem has two choices: either the problem is ascribed to some physical process affecting the measuring rods, such as the heating of the table, or it is attributed to the geometry of the table. If one has other measuring devices that are not subject to deformation due to heating, the problem can easily be settled. But if all measuring devices behave in the same way, the ambiguity remains of whether one should assume that Euclidean geometry nevertheless holds, although it cannot be directly established due to the intervention of some universal force affecting all measurements, or whether one should admit the possibility of non-Euclidean geometry.

Without the distinction between axiomatic Euclidean geometry and practical rigid-body geometry, we arrive at the view advocated by Poincaré.[15] He suggested that the choice of geometry to describe spatial properties is ultimately conventional. For the conventionalist, the axioms of geometry are neither analytic nor synthetic truths in the sense of Kant. They are not necessary truths either. They are definitions. When we find entities that satisfy these definitions, the axioms become true of these entities. The only constraint on axioms is consistency. Since we do not know in advance, by intuition, say, that the axioms are true, proving their consistency is crucial. Poincaré thought that the combination of geometry and physics does indeed apply to experience. He was convinced,

though, that there could be different combinations of geometry and physics that would satisfy the empirical criteria. But since Euclidean geometry is the simplest of all alternative geometries, it will always be adopted as the most natural geometrical component of such a combination.

For Einstein, however, measuring rods retain their foundational role for defining the geometrical properties of space even if they may be subject to deformations, at least as long as rods of all materials suffer the same deformation. For him, there is no ambiguity because geometry itself is an empirical science founded in what he called practical geometry. Of course, it would be desirable that such practical geometry would ultimately be derived from a more comprehensive theory so that Einstein's stance may be viewed as a preliminary one, shaped by a particular historical situation, to be improved upon once a theory describing the measurement instruments becomes available. This issue was raised in a discussion with Hermann Weyl in the context of his proposed extension of Riemannian geometry (chapter 8, p. 126). As was emphasized above, Einstein's interpretation of measuring instruments in terms of practical geometry, as preliminary as this may have been, was crucial for his formulation of general relativity.[16] Had he instead accepted the conventionalist position, he might have tried to attribute any deviant behavior of measuring instruments to some speculative physical processes rather than to a change of the concepts of space and time. In holding onto the meaning of these instruments given by "practical geometry," Einstein was prevented from pursuing such speculations and opened up the possibility of drawing consequences for understanding space and time from their behavior, as strange as they might appear.

This in fact is exactly the situation in which Einstein introduced the mental model of a rotating disk, which he used as early as 1912 to show that the new theory of ravitation requires a new framework for space and time. An observer on the disk is equipped with standard clocks and rods. But using these measuring instruments to coordinate space and time measurements within the rotating frame of reference turns out to be difficult when the implications of special relativity are taken into account. A clock at the rim of the rotating disk and a clock sitting at its center will not tick at the same pace because they live in reference frames moving with respect to one other. Pursuing the same thought experiment with rods, Einstein derives the even more far-reaching consequence that for the observer on the rotating disk Euclidean geometry no longer holds because the behavior of the standard rods also depends on the place at which they are used. Suppose that an observer moving along with the rotating disk uses standard rods to measure the circumference of the disk. According to special relativity, these rods are shortened with respect to an observer at rest (Lorentz contraction), therefore more rods are needed and the length of the circumference exceeds the result of such a measurement for a nonrotating disk. At the same time, the length of the rods used to measure the diameter is unaffected. Therefore, the ratio of the circumference to the diameter is greater than π. Thus, Euclidean geometry does not apply to the rotating system.[17]

Einstein presented the rotating-disk argument in his introductory remarks to general relativity in the third Princeton lecture. The conclusion of this argument is (in Einstein's own words):

> Space and time, therefore, cannot be defined with respect to K′ (reference frame of the disk) as they were in the special theory of relativity with respect to inertial systems. But according to the principle of equivalence, K' may also be considered as a

> system at rest, with respect to which there is a gravitational field (field of centrifugal force, and force of Coriolis). We therefore arrive at the result: the gravitational field influences and even determines the metrical law of space-time continuum. If the laws of configuration of ideal rigid bodies are to be expressed geometrically, then in the presence of a gravitational field the geometry is not Euclidean.[18]

Gravitation here plays the role of heating in Poincaré's marble table illustration.

The *New York Evening Post* of 12 May reported that at the colloquium convened to discuss the difficulties of the fourth lecture (the third in the printed version), the principal point of discussion was related to the rotating disk. "Prof. Einstein removed the difficulties which had been felt by remarking that the objective existence of the rotating disk was not at all necessary for the purpose of the illustration."

Einstein had previously referred to this difficulty. In a copy of the tenth German edition of *Relativity—The Special and General Theory*, which was found at the Harry Ransom Humanities Research Center at the University of Texas, there is a sheet of paper with a remark in the handwriting of Ilse Einstein, the daughter of Einstein's wife Elsa, who at that time served as his secretary. The remark refers to the discussion of the rotating disk. It reads:

> Note: This interpretation has often been opposed as unconvincing because not only the measuring rods but also the circular disk would suffer tangential contraction. This argument is not cogent because the rotating disk cannot be viewed as a Euclidean rigid body; such a body would shatter when brought to rotation, according to the postulated reasons. In reality the disk plays no role in the whole consideration, but only the system—of rods at rest relative to each other—which rotates as a totality of radially and tangentially positioned rodlets.

NOTES

1. For a discussion of "Geometry and Experience," see Michael Friedman, "Space, Time, and Geometry," in *The Cambridge Companion to Einstein*, ed. M. Janssen and C. Lehner (Cambridge: Cambridge University Press, 2014), pp. 398–420.
2. Gutfreund and Renn, *Relativity: The Special and the General Theory*.
3. Albert Einstein, "Geometry and Experience," in *Sidelights of Relativity*, ed. A. Einstein (London: Methuen, 1922), pp. 17–18.
4. Abraham Pais, *'Subtle is the Lord . . .': The Science and Life of Albert Einstein* (Oxford: Oxford University Press, 1982), p. 134.
5. See also chapter 7.
6. See Jimena Canales, *The Physicist and the Philosopher: Einstein, Bergson, and the Debate That Changed Our Understanding of Time* (Princeton, NJ: Princeton University Press, 2015). See also chapter 7, p. 109.
7. "Manuscript on the Special Theory of Relativity," CPAE vol. 4, Doc. 1.
8. David E. Rowe, "Max von Laue's Role in the Relativity Revolution," *Mathematical Intelligencer* 30 (September 2009): 54–60.
9. Albert Einstein, "Die Grundlage der allgemeinen Relativitätstheorie," *Annalen der Physik* 49 (1916): 769–822. English translation in CPAE vol. 6, Doc. 30, p. 146.
10. See "The *Entwurf* Theory," p. 42.
11. Albert Einstein, "The Formal Foundation of the General Theory of Relativity," CPAE vol. 6, Doc. 9.

12. Janssen and Renn, "Arch and Scaffold."
13. Albert Einstein, 1 May 1935, "The Late Emmy Noether: Professor Einstein Writes in Appreciation of a Fellow-Mathematician," in "Letters to the Editor," *New York Times*, 4 May 1935, p. 12.
14. See Gutfreund and Renn, *Relativity: The Special and the General Theory*.
15. For an extensive analysis of Poincaré's conventionalism, see Yemima Ben-Menahem, *Conventionalism* (Cambridge: Cambridge University Press, 2006).
16. For extensive discussions, see Alexander Blum, Jürgen Renn and Matthias Schemmel, "Experience and Representation in Modern Physics: The Reshaping of Space," in *Spatial Thinking and External Representation: Towards a Historical Epistemology of Space*, ed. M. Schemmel (Berlin: Edition Open Access, 2016), http://www.edition-open-access.de/studies/8/index.html.
17. For more details on this argument, see Michel Janssen, "'No Success Like Failure': Einstein's Quest for General Relativity, 1907–1920," in *The Cambridge Companion to Einstein*, pp. 167–227, p. 181.
18. CPAE vol. 7, Doc. 71, pp. 320–321.

2
THE PRINCIPLES OF GENERAL RELATIVITY

What did Einstein want to achieve when he embarked on his intellectual journey toward a theory of general relativity, and what did he actually achieve? He clearly wanted to generalize the principle of relativity to apply to any type of motion. In his popular account of the special and general theories of relativity, Einstein stated:

> Since the introduction of the special principle of relativity has been justified, every intellect which strives after generalization must feel the temptation to venture the step towards the general principle of relativity.[1]

In the seminal review article "The Foundation of the General Theory of Relativity," he explained how Mach's criticism of Newton's mechanics led him to this goal.[2]

A. THE ORIGIN OF THE GENERALIZATION OF THE RELATIVITY PRINCIPLE

Newtonian mechanics is based on the notion that space is absolute by its nature—without relation to extraneous objects, always the same and immovable—and that there is a distinction between relative and absolute motion. Isaac Newton demonstrated this distinction using his famous bucket experiment (figure 4). One of two identical buckets filled with water is rotating. The parabolic surface of water in the rotating bucket is caused by the centrifugal force, which is present only in this bucket. The centrifugal force is an example of the fictitious (inertial) forces that appear in accelerated frames of reference.

According to Newton, the rotation occurs with respect to absolute space, and the presence of the centrifugal force distinguishes between absolute and relative motion. Mach rejected the notion of absolute space and absolute motion, suggesting that the shape of the water surface in the rotating bucket can be explained as being produced by rotation with respect to the rest of the matter in the universe. Mach conjectured that if the bucket were at rest and all the fixed stars in the universe were rotating, the same parabolic shape of the water surface would have been formed.

Einstein understood well, of course, that there is a difference between uniform and accelerated motion. A person riding on a train can sense the difficulty of such a generalization. If a sudden acceleration is applied to the train, that person will be subject to a tilt

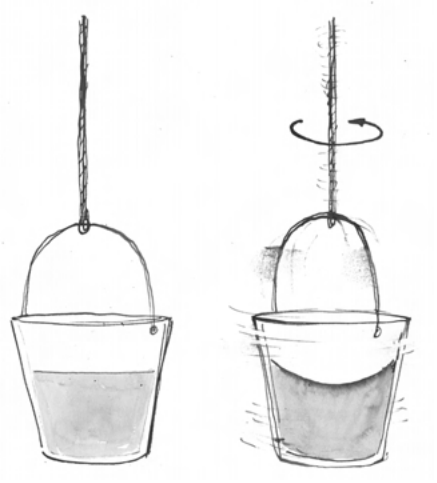


FIGURE 4: Einstein's version of Newton's celebrated "two-bucket experiment." (Laurent Taudin)

backward or forward. Thus, the mechanical behavior of bodies under such circumstances is different from the behavior in a Galilean (inertial) system of coordinates. It is then only natural to attribute an absolute physical reality to accelerated motion. But Einstein nevertheless managed to argue that accelerated frames of reference should be admitted on an equal footing with the familiar inertial frames of reference from classical physics and special relativity, if one is allowed to introduce, at the same time, the idea of a relativity of the gravitational field. It is here that the equivalence principle comes in, which Einstein later characterized as the happiest thought of his life.[3]

Indeed, in 1907 the equivalence principle had helped him to overcome a puzzling problem in his attempts to integrate the phenomenon of gravitation into the framework of the special theory of relativity. Such an attempt seemed to clash with the well-established principle of classical physics that all bodies fall with the same acceleration. This suggested that he simulate a constant gravitational field by introducing a reference frame in uniform acceleration in which, for kinematic reasons, all bodies would have to fall with the same acceleration. The idea of considering such an accelerated system as a useful device to explore gravity in a relativistic context resonated with several other ideas on his mind, which must have been one of the reasons why he found this thought so striking. One of these ideas was Mach's criticism of Newton's concept of absolute space, suggesting a generalization of the relativity principle to all motions, including accelerated motions. In this sense, the uniformly accelerated frame of reference introduced by Einstein and later often described in terms of the *elevator thought experiment* was nothing but a simplified version of the bucket thought experiment of Newton and Mach. Another idea that must have struck Einstein was the analogy of this situation with the dependence of the manifestation of the electromagnetic field on relative motion. Indeed, in 1920, in his essay "Fundamental Ideas and Methods of the Theory of Relativity Presented in their Development," Einstein emphasized this close analogy between the origins of special and general relativity:

> In an example worth considering, the gravitational field only has a relative existence only in a manner similar to the electric field generated by magneto-electric induction.

> *Because for an observer in free-fall from the roof of a house, there is during the fall*—at least in his immediate vicinity—*no gravitational field*. Namely, if the observer lets go of any bodies, they remain relative to him, in a state of rest or uniform motion, independent of their special chemical or physical nature. The observer, therefore, is justified in interpreting his state as being "at rest."[4]

The equivalence principle thus embodied both the idea of a generalization of the relativity principle to accelerated motions and the relative existence of a gravitational field. Only if this relative existence is admitted does the idea of a generalization of the relativity principle make sense. The equivalence principle quickly became the most important heuristic tool in creating general relativity.

On closer inspection, the equivalence principle represents a hybrid principle that combines ideas of mechanics, such as the relativity of motion, with ideas from field theory, such as the relativity of the gravitational field.[5] In contemporary physics, the traditions of mechanics and field theory were pursued separately from each other, in particular with regard to the problems at hand. The problem of a generalization of the relativity principle was also addressed, for instance, by some of Mach's followers, who tried to explain the inertial forces occurring in an accelerating frame by an as-yet-unknown interaction between distant masses in relative motion with respect to each other. They searched in vain for direct empirical evidence of such a mysterious interaction but did not realize how to relate it to gravitation.

Physicists within the tradition of field theory, on the other hand, searched for empirical clues to guide them beyond Coulomb's law of static gravitation (that is, Newton's law) to a full gravitational field theory, but they did not consider inertial forces to be providing just such a clue. Einstein's position was rather unique in that he was both a champion of field theory and a fervent adherent to Mach's ideas about inertia. This position allowed him to formulate the equivalence principle as a starting point for a field theory of gravitation in which inertial forces could provide special cases for the otherwise unknown dynamical gravitational fields.

Einstein had thus found a "third way" and the key to the problems that appeared to be unsolvable within each of the two traditions—mechanics and field theory—taken separately.[6] But he had also created a tension between the notion of a general principle of relativity and the relativity of the split of the unified gravito-inertial field into a gravitational and an inertial component, a tension that was to shape discussions during the formative years.

B. THE GROWING TENSION BETWEEN EINSTEIN'S HEURISTICS AND HIS ACCOMPLISHMENTS

In his early publications on general relativity, Einstein insisted on its epistemological advantages, which provided additional support for its claim of superiority with regard to competing theories. As will be discussed in greater detail in chapter 5, for some time between 1916 and early 1917, Einstein attempted to formulate boundary conditions for his field equations that would comply with his Machian heuristics, according to which inertial effects should be due to the interaction of masses. After the failure of these attempts, he introduced, in his famous 1917 cosmological paper, a spacetime that satisfied all his expectations concerning the explanation of its inertial properties by the masses acting as sources of the gravitational field, but this was done at the price of modifying the field equations for which this spacetime was a solution. It was therefore an unpleasant surprise

when he realized that even the modified field equations did not constitute a Machian theory, in the sense that they also had solutions describing a non-Machian spacetime in which inertial effects are not necessarily caused by masses. As a result, Einstein's expectation that general relativity was to provide a Machian explanation of inertia began to be silently transformed from a requirement concerning the nature of the theory to a criterion to be applied to specific solutions of the theory.

The increasing tension between Einstein's original intentions and the ongoing exploration of the consequences of the new theory was enhanced when, in 1917, Erich Kretschmann challenged Einstein's interpretation of the meaning of general covariance.[7] Einstein had identified his general principle of relativity with general covariance:

> As all our physical experience can be ultimately reduced to such coincidences, there is no immediate reason for preferring certain systems of coordinates to others, that is to say, we arrive at the requirement of general covariance.[8]

Using the same concept of point coincidences, Kretschmann argued that general covariance has no physical meaning and that with mathematical ingenuity any spacetime theory, with or without absolute motion, can be expressed in a generally covariant formulation. He concluded this from the assertion that if

> all physical observations consist in the determination of purely topological relations ("coincidences") between objects of spatio-temporal perception, from which it follows that no coordinate system is privileged by these observations, then one is forced to the following conclusion: By means of a purely mathematical reformulation of the equations representing the theory, . . . any physical theory can be brought into agreement with any, arbitrary relativity postulate, even the most general one.[9]

With this argument, Kretschmann challenged the close link between covariance and the relativity principle.

C. THE PRINCIPLES OF GENERAL RELATIVITY REVISITED

In 1918, Einstein reacted to Kretschmann's criticism and also drew the consequences from the problems with implementing the Machian interpretation of general relativity mentioned above. He argued that he had so far not sufficiently distinguished between two principles that he now introduced as the principle of relativity and Mach's principle. Einstein formulated his understanding of the underlying principles of the theory of general relativity as follows:

1. *Principle of Relativity*. Nature's laws are merely statements about temporal-spatial coincidences; therefore they find their only natural expression in generally covariant equations.
2. *Principle of Equivalence*. Inertia and gravity are phenomena identical in nature. From this and from the special theory of relativity it follows necessarily that the symmetric "fundamental tensor" ($g_{\mu\nu}$) determines the metric properties of space, the inertial behavior of bodies in this space, as well as the gravitational effects. We shall call the state of space, which is described by this fundamental tensor the "*G*-field."

3. *Mach's Principle*. The *G*-field is completely determined by the masses of the bodies. Since mass and energy are—according to the results of the special theory of relativity—the same, and since energy is formally described by the symmetric energy tensor ($T_{\mu\nu}$), it follows that the *G*-field is caused and determined by the energy tensor of matter.[10]

The relativity principle is not formulated here, as was done previously, as the equivalence of all frames of reference. Instead, Einstein states that the only physically meaningful content of a relativistic theory are coincidences of physical events at points of space and time. Since the occurrence of these point coincidences is independent of whether they are described in one or the other coordinate frame, their most appropriate description is by a generally covariant theory. This principle, of course, had not been the starting point of Einstein's search for a generally relativistic theory of gravitation, but instead constitutes a result of his reflection on complications encountered in a long but eventually successful search for such a theory. The formulation of the principle of relativity no longer appeals to the original intuition of a world of isolated bodies distributed in an otherwise empty space whose physical interactions should depend only on their relative distances, velocities, and so on.

Einstein agreed with Kretschmann that this formulation of the relativity principle is not a statement about physical reality but instead is a requirement about the mathematical formulation of the theory, and he even conceded that "every empirical law can be brought into a generally covariant form." However, he nevertheless attached heuristic value to principle 1, which has proven itself splendidly in the relativistic theory of gravitation. Einstein claimed that between two formulations of the same theory, one will always favor the simpler one. It later turned out that it was actually possible to formulate simple covariant versions of Newton's theory.[11] But in contrast to general relativity, such a theory presupposes absolute spacetime structures that may act but are not acted upon by physical processes, which made it clear that the characteristic feature of Einstein's theory can be best captured by the notion of background independence, in the sense that the theory does not presuppose such absolute spacetime structures.[12]

The equivalence principle was formulated before as enabling the generalization of the relativity principle from uniform to accelerated motion. It was the starting point of the whole theory and led to principle 1. Here the formulation is different. The emphasis is on the equivalence of inertia and gravity. In the third Princeton lecture (pp. 217–18 [57–58]), the equivalence principle is formulated in the usual fashion. As he had done many times before, Einstein introduced a coordinate system, *K'*, which is uniformly accelerating with respect to an inertial coordinate system *K*. Thus:

> Relatively to *K'* all the masses have equal and parallel accelerations; with respect to *K'* they behave just as if a gravitational field were present and *K'* were unaccelerated. (p. 217 [57])

In this situation

> there is nothing to prevent our conceiving this gravitational field as real, that is, the conception that *K'* is "at rest" and a gravitational field is present we may consider as equivalent to the conception that only *K* is an "allowable" system of co-ordinates and

> no gravitational field is present. The assumption of the complete physical equivalence of the systems of coordinates, *K* and *K'*, we call the "principle of equivalence;" this principle is evidently intimately connected with the law of the equality between the inert and the gravitational mass, and signifies an extension of the principle of relativity of co-ordinate systems which are in non-uniform motion relatively to each other. In fact, through this conception we arrive at the unity of the nature of inertia and gravitation. (pp. 217–18 [57–58])

Only the last sentence reflects the essence of the formulation of the equivalence principle in Einstein's response to Kretschmann. Einstein then gives an almost poetic expression to this point:

> The possibility of explaining the numerical equality of inertia and gravitation by the unity of their nature gives to the general theory of relativity, according to my conviction, such a superiority over the conceptions of classical mechanics, that all the difficulties encountered must be considered as small compared with this progress. (p. 218 [58])

As for Mach's principle, Einstein commented in 1918:

> The necessity to uphold it is by no means shared by all colleagues; but I myself feel it is absolutely necessary to satisfy it. With (c)[i.e., Mach's principle], according to the field equations of gravitation, there can be no *G*-field without matter. Obviously, postulate (c) is closely connected to the space-time structure of the world as a whole, because all masses in the universe will partake in the generation of the *G*-field.[13]

In this way Einstein had translated Mach's ideas from the language of mechanics to that of field theory because both terms in Einstein's 1918 definition of Mach's principle—the gravitational field and the energy-momentum tensor—are field theoretical concepts.

Subsequently, Einstein extended the field theoretical interpretation of general relativity at the expense of the emphasis on the mechanical roots of his original heuristics. By 1920, the attempt to define Mach's principle in terms of the field-theoretical building blocks of his theory had been complemented by the introduction of a "Machian ether" as a means of capturing its conceptual implications. In a lecture given in Leiden,[14] Einstein exploited the time-honored concept of an ether, to which Lorentz had given the definitive form in the realm of electrodynamics in order to explain the new concept of spacetime, which had emerged with general relativity. Einstein now directly turned against Mach's interpretation of inertial effects as caused by cosmic masses, because this interpretation presupposed an action at a distance, a notion incompatible with both field theory and relativity theory. Instead, contrary to his original heuristics, Einstein associated these inertial effects with the nature of space, which he now conceived as being equipped with physical qualities and which he hence appropriately called ether. Contrary to Lorentz's ether, however, Mach's ether, which Einstein thought of as being represented by the metric tensor, was not only supposed to condition but also to be conditioned, at least in part, by matter. This capacity of being influenced by the presence of matter was apparently the last resort for the Machian idea of the generation of inertial effects by the interaction of material bodies in Einstein's conceptual framework.[15]

In the course of his work on unified field theory, which is discussed in chapter 8, Einstein even came to favor a causal primacy of space in relation to matter—in sharp contrast to his original Machian heuristics. In a lecture given in 1930, Einstein formulated this view drastically:

> The strange conclusion to which we have come is this—that now it appears that space will have to be regarded as a primary thing and that matter is derived from it, so to speak, as a secondary result. Space is now turning around and eating up matter. We have always regarded matter as a primary thing and space as a secondary result. Space is now having its revenge, so to speak, and is eating up matter.[16]

When the occasion presented itself much later, Einstein also became quite explicit about his rejection of his earlier Machian heuristics. In a letter to the physicist Felix Pirani, for instance, he explains with reference to Mach's principle, as he himself had earlier defined it, that he no longer finds it plausible that matter represented by the energy-momentum tensor could completely determine the gravitational field, since the specification of the energy-momentum tensor itself already presupposes knowledge of the metric field. In the same letter Einstein explicitly revokes Mach's principle:

> In my view one should no longer speak of Mach's principle at all. It dates back to the time in which one thought that the "ponderable bodies" are the only physically real entities and that all elements of the theory that are not completely determined by them should be avoided. (I am well aware of the fact that I myself was long influenced by this idée fixe.)[17]

D. THE RESILIENCE OF THE RELATIVITY PRINCIPLE

In spite of having agreed with Kretschmann's criticism, in *The Meaning of Relativity* (third Princeton lecture, p. 221 [61]) Einstein quickly returned to his original formulation of the link between general covariance and the relativity principle:

> We shall be true to the principle of relativity in its broadest sense if we give such a form to the laws that they are valid in every such four-dimensional system of co-ordinates, that is, if the equations expressing these laws are co-variant with respect to arbitrary transformations. (p. 221 [61])

Why did he find the relativity principle so irresistible? Was it because it constituted the most central of his heuristic starting points? But as we have just seen, he was then ready to modify his position on practically all other heuristic principles that had originally guided him. To answer this puzzling question, we have to once more turn to Mach's reinterpretation of Newton's thought experiment, trying to understand the irresistible plausibility of the relativity principle.

In "The Foundation of the General Theory of Relativity," Einstein described his own version of Newton's bucket argument, which was probably also inspired by Newton, who had considered as another example two globes rotating around each other, with the

tension in the cord that held them together revealing their acceleration.[18] Einstein also referred to a system of two fluid bodies of the same size and nature separated in space. One of them is rotating at constant angular velocity around the axis connecting the two bodies. An observer at rest with respect to either body sees the other one as rotating, yet the shape of only one of them is changed as a result of this rotation. Einstein could not see any possible cause within the system of the two bodies that could be responsible for their different behavior. Following Mach, he therefore concluded that it must be attributed to a cause outside the system, namely, the distant masses in the universe. None of the reference frames with respect to which one of the two bodies is at rest seemed to be privileged. This conclusion led Einstein to the general principle of relativity: "The laws of physics must be of such a nature that they apply to systems of reference in any kind of motion."[19]

In the Princeton lectures, he not only subscribed to the same principle in spite of Kretschmann's criticism, but he made it even more explicit that his intuition was shaped by the very concept of motion, as he explained at the beginning of his first popular lecture:

> When we generally speak of the motion of a body, we always mean relative motion by the concept of motion. For example, when we speak of a car moving in the street, the motion refers to the piece of ground or surface called the road, and this piece of earth's surface plays the role of a body on which this motion will develop. Thus motion is conceived as relative motion, and according to this concept one could equally well say the street moves relative to the car, as we can say the car moves relative to the street. These are quite self-evident circumstances, and in a sense have nothing to do with relativity theory. Anyway the fact is that we can conceive of motion only as relative motion. Concerning the pure geometrical acceleration it is indifferent, from which body [one talks about it]. That is self-evident and needs no detailed elaboration.[20]

In his contribution to *The Cambridge Companion to Einstein*, the historian Michel Janssen begins with this quotation and remarks that this is not only not self-evident, but not even true.[21]

In an unpublished manuscript, which Michel Janssen believes to be an early draft of the published version of the Princeton lectures, Einstein explicitly introduced a distinction between what he called "kinematical" and "physical" relativity: [22]

> *Kinematical relativity of motion*. That motion by definition can only be conceived of as relative motion of one body with respect to another is a fact that has long been known to philosophers. One can also express this as follows: from the point of view of pure kinematics there is only relative motion. . . .
>
> *Physical relativity of motion*. This is by no means automatic: Even though it is a fact that all states of motion are equivalent from a purely kinematical point of view, this equivalence need not exist at all from a physical point of view. . . .
>
> *Physical relativity of motion is the fundamental assumption of relativity theory*: . . . The general theory of relativity is based on the presupposition that there are no physically privileged states of motion in nature at all, and that, therefore, all coordinate systems are equivalent for the formulation of the laws of nature.

Einstein evidently took it for granted that at least kinematical relativity holds for the situation described in his version of the bucket experiment. This is indeed the case if this situation

is conceptualized in terms of classical Newtonian physics, which evidently continued to shape Einstein's intuition about it. On closer inspection, however, Einstein's assessment of the situation suffers from a number of rather subtle shortcomings, which partly became a subject of discussion during the formative years and were partly clarified only much later. We thus witness that the knowledge involved in such thought experiments has a layered structure: one layer being shaped by common sense, or in this case Newtonian intuitions, and another layer in which the technical details of the theories at issue matter.

Indeed, in a relativistic context, the kinematical equivalence of the two reference systems, so plausible according to the Newtonian intuition, becomes a more intricate issue. For instance, in classical physics there is no problem with switching from a reference frame in which a body is rotating with respect to the fixed stars to a reference frame in which instead the body is at rest and the fixed stars are rotating, whatever the speed of rotation is; but if the stars have to rotate with superluminal velocity in order to maintain the equivalence, this is obviously a problem in a relativistic setting. Also, in classical physics, such a kinematic comparison does not raise issues about the synchronization of clocks, whereas in a relativistic context it does.

A generally relativistic response to the bucket experiment would entail that the centrifugal forces acting on the water in the bucket could also be traced to the rotation of ponderable masses around a bucket at rest. Einstein had pursued this line of thinking since the beginning of his work on a relativistic theory of gravitation. As early as 1912, he had considered a giant rotating spherical shell of matter, representing the fixed stars in Mach's thought experiment, which was supposed to generate near its center a gravitational field corresponding to that of a rotating frame of reference. He thus attempted to establish the concept of *rotation as rest*.[23] The failure to implement this idea in the preliminary *Entwurf* theory (see box) represented a major blow for this theory. But even according to the final version of general relativity, this is not the case since the field produced by the rotating shell is much weaker than that occurring in a rotating reference frame. Einstein had recognized this failure by the time of his Princeton lectures. In his text, he referred to the Coriolis and centrifugal field inside a hollow body but admitted that their effects are too small to be detected. Nevertheless, he still included this result in the list of arguments in favor of Mach's principle (p. 260 [100]).

THE *ENTWURF* THEORY

In the spring of 1913, Einstein derived a gravitational field equation that became known as the core of the *Entwurf* theory.[24] It primarily satisfied principles rooted in classical physics, namely, energy-momentum conservation, and it reduces to Newton's theory in the limit of weak static gravitational fields. To satisfy these requirements, as he then understood them, he had to impose certain restrictions on admissible coordinates. As a result, the class of coordinate systems in which the *Entwurf* equation takes on the same form does not satisfy the generalized principle of relativity in the way he imagined. He therefore abandoned with a heavy heart the realization of general covariance. Nevertheless, he could reassure himself that this equation was acceptable because the necessary restriction of the admissible coordinate systems could apparently be justified by the requirement to implement

the conservation principle. From the modern perspective, this theory is incorrect, but at that time Einstein assumed that it was the best that could be done, and he even convinced himself that a gravitational field theory does not allow general covariance. In late 1915, Einstein realized the fallacy of this conclusion. Revisiting the *Entwurf* theory led him, in a short period of time, to his final formulation of the general theory of relativity.

But even after this failure of a strong version of the relativity of motion, one might still expect that at least a weaker version holds. This would entail that the outward bulge in the globe in Einstein's variant of the bucket experiment can be explained on the basis of the same physical laws in both reference systems, the one in which the stars are at rest and the one in which the globe is at rest, without requiring the rotation of the stars to produce the same effect as the rotation of the globe. This explanation, however, is also not as obvious as it appeared to Einstein, as later commentators have noted.[25] In particular, if two observers—one moving with the rotating body and the other at rest—use clocks, the first observer would find that a clock on the surface of the rotating body does not keep the same time as a clock at rest. Also, a description of the rotating body in a rotating frame and its description in a nonrotating frame would not satisfy the same boundary conditions. But if the kinematical equivalence of the two situations to be compared can no longer be taken for granted, observers who find that their coordinate systems are not equivalent will realize that an effect like the bulging of the globe can, after all, be attributed to a particular state of motion, thus challenging a general principle of relativity.

Of course, in a generally covariant theory, observers may use any coordinate system. In general relativity, all systems of coordinates are, in principle, admissible. For Einstein, as we have seen, general covariance was the mathematical expression of the general principle of relativity:

> That this requirement of general co-variance, which takes away from space and time the last remainder of physical objectivity, is a natural one, will be seen from the following reflection. All our space-time verifications invariably amount to a determination of space-time coincidences.[26]

However, Einstein's thought experiments require that coordinate systems retain just enough spatiotemporal meaning to define and compare motions. Otherwise they would no longer suggest that one should account for kinematically equivalent physical situations with dynamically equivalent explanations. But as it turns out, this kinematical equivalence was founded on a classical intuition that Einstein never abandoned because he was convinced that it was part of the very concept of motion as expressed in what he called the kinematical relativity of motion. In another contemporary unpublished manuscript, possibly also directly related to the Princeton lectures, Einstein wrote:

> It has been clear for ages that motion can only be thought of as relative motion (motion of one body with respect to another), not as absolute motion (motion of a body not referred to other bodies). It is therefore impossible, as long as one bases oneself solely on the concept of motion, to prefer one state of motion over all other states of motion through some special marks.[27]

However, because of the difficulties in establishing such a kinematical equivalence for his thought experiments under relativistic conditions, from today's point of view it is hardly possible to subscribe to Einstein's claim that the general principle of relativity captures the essence of his theory. What remains instead is his recognition of the relativity of the gravitational field, that is, of the dependence of the split of the gravito-inertial field into gravitational and inertial components on the observer's state of motion, as entailed by the equivalence principle.[28] This reinterpretation of one of Einstein's key insights illustrates a more general phenomenon, the fact that the elaboration of general relativity during the formative years and beyond eventually superseded Einstein's heuristics, which had formed the very starting point of the entire endeavor. Theory change is hardly ever exclusively the matter of a momentary paradigm shift due to an individual achievement, but instead is typically a long-term and cooperative effort.[29]

To summarize, the formative years were marked by debates about the two main heuristic principles that had guided Einstein on the road to his relativistic theory of gravity: the general principle of relativity and Mach's principle. In the course of these debates, both principles were challenged and substantially modified. The eventual demise of Mach's principle in connection with the rise of relativistic cosmology and the prevalence of field theoretical concepts is well known; less so the problematic character of the general principle of relativity. Even today we can find statements that the relativity of motion is the earmark of the theory of general relativity. As we have seen, this is largely due to the intuitive plausibility of kinematic relativity on the basis of a classical understanding of motion that is actually no longer tenable according to Einstein's own theory. Remarkably, some of its more subtle conceptual implications were only recognized in the context of careful analyses pursued in recent collaborations of philosophers, historians, physicists, and mathematicians.[30]

NOTES

1. Albert Einstein, *Relativity: The Special and General Theory*, trans. R.W. Lawson (London: Methuen, 1920), 56. See also Gutfreund and Renn, *Relativity: The Special and the General Theory*.
2. CPAE vol. 6, Doc. 30. See also Gutfreund and Renn, *Road to Relativity.*
3. Gutfreund and Renn, *Road to Relativity*, p. 45.
4. CPAE, vol. 7, Doc. 31, p. 136.
5. For an amplification of some of the arguments of one of his teachers, Jon Dorling, see Michel Janssen, "The Twins and the Bucket: How Einstein Made Gravity Rather Than Motion Relative in General Relativity," preprint, *PhilSci Archive*, 2011, http://philsci-archive.pitt.edu/8605/.
6. See Jürgen Renn, "The Third Way to General Relativity," in *The Genesis of General Relativity*, vol. 3, ed. J. Renn and M. Schemmel (Dordrecht: Springer, 2007), pp. 21–75.
7. For a discussion of Kretschmann's criticism and Einstein's response, see J. D. Norton, "General Covariance and the Foundations of General Relativity: Eight Decades of Dispute," *Reports on Progress in Physics* 56 (1993): 791–858.
8. Albert Einstein, "The Foundation of General Theory of Relativity," CPAE vol. 6, Doc. 30, p. 154.
9. Erich Kretschman, "Über den physikalischen Sinn der Relativitätspostulate, A. Einsteins neue und seine ursprüngliche Relativitätstheorie," *Annalen der Physik* 358 (1918): 575–614. Translation from John Norton, "Did Einstein Stumble? The Debate over General Covariance," in *Reflections on Spacetime: Foundations, Philosophy, History*, ed. U. Majer and H.-J. Schmidt (Dordrecht: Kluwer Academic Publishers, 1995), pp. 103–125, p. 108.
10. CPAE vol. 7, Doc. 4, pp. 33–34.

11. See Èlie Cartan, "Sur les varietes a connexion affine et la theorie de la relativitee generalisee (premiere partie)," *Annales scientific de l'E.N.S.* 3 (1923): 325–412.
12. John Stachel, "The Hole Argument and Some Physical and Philosophical Implications," *Living Rev. Relativity* 17 (2014); Ben-Menahem, *Conventionalism*.
13. CPAE vol. 7, Doc. 4, p. 34.
14. Albert Einstein, "Ether and the Theory of Relativity," CPAE vol. 7, Doc. 38, p. 163.
15. For historical reviews of the role of Mach's principle in Einstein's thinking, see Renn, "'The Third Way to General Relativity"; Carl Hoeffer, "Mach's Principle as Action-at-a-Distance in GR: The Causality Question," *Studies in History and Philosophy of Modern Physics*, 48 B (November 2014), pp. 28–136; Janssen, " 'No Success Like Failure.' "
16. Albert Einstein, "Professor Einstein's Address at the University of Nottingham," *Science* 71 (1930): 608–610, p. 610. (Address given in German and translated by Dr. I. H. Brose.)
17. Einstein to Felix Pirani, 2 February 1954, Einstein Archives, call no. 17-447.00. Authors' translation.
18. Janssen, " 'No Success Like Failure.' "
19. CPAE 6, Doc. 30, p. 149.
20. Translation by D. Brill. The phrase in the square brackets is our interpretation of what is missing in the transcript of the stenographer's text.
21. Janssen, " 'No Success like Failure.' "
22. "On the Special and General Theory of Relativity," first lecture, section 1, "Relativity of Motion and the Relativity Principle," §§1–3, CPAE vol. 7, Doc. 63, pp. 449–450. See Janssen, "Twins and the Bucket."
23. Albert Einstein, "Is There a Gravitational Effect Which Is Analogous to Electrodynamic Induction?" *Vierteljahrsschrift fur gerichtliche Medizin und öffentliches Sanitätswesen* 44 (1912): 37–40.
24. Gutfreund and Renn, *Road to Relativity*, pp. 24–26.
25. Jon Dorling, "Did Einstein Need General Relativity to Solve the Problem of Absolute Space? Or Had the Problem Already Been Solved by Special Relativity?" *British Journal for the Philosophy of Science* 29 (1978): 311–323.
26. Einstein, "Foundation of General Relativity," CPAE vol. 6, Doc. 30, p. 153.
27. "Lecture on the Special Theory of Relativity," first lecture [p. 2], second and third full paragraphs, CPAE, Vol. 7, Doc. 64, pp. 459–460.
28. Janssen, "Twins and the Bucket.".
29. See Jürgen Renn and Tilman Sauer, "Pathways out of Classical Physics: Einstein's Double Strategy in Searching for the Gravitational Field Equation," in *The Genesis of General Relativity*, vol. 1, Michel Janssen, Jürgen Renn, John D. Norton, Tilman Sauer, and John Stachel (Dordrecht: Springer, 2007), chap. 8, pp. 292–306; and Janssen, " 'No Success Like Failure.' "
30. See Dorling, "Did Einstein Need General Relativity," and Janssen, "Twins and the Bucket."

3
THE FIRST SOLUTIONS AND THE CHALLENGE OF THEIR INTERPRETATION

When Einstein published the final form of the field equations on 25 November 1915, only an approximate solution was known. He had first developed an approximation scheme in the framework of the *Entwurf* theory, then adapted it to the final version and used it to successfully calculate the perihelion motion of Mercury. Given the complicated nonlinear character of the field equations, he did not expect that exact solutions could easily be found. He was all the more surprised when the astronomer Karl Schwarzschild presented him with just such an exact solution. On 9 January 1916, Einstein wrote to him:

> I have examined your paper with great interest. I would not have expected that the exact solution of the problem could be formulated so simply. The mathematical treatment of the subject appeals to me exceedingly. Next Thursday I am going to deliver the paper before the Academy with a few words of explanation.[1]

A. THE SCHWARZSCHILD SOLUTION

Schwarzschild's exact solution represents the field of a spherically symmetric mass distribution such as that of the sun. Schwarzschild began his approach by imposing the coordinate condition that requires the absolute value of the determinant of the metric to be one. This determinant condition governed Einstein's treatment of late 1915, albeit no longer for reasons of principle as in his earlier versions of the theory. Nevertheless, this choice of coordinates left a trace in some of the subsequent work. In the case of Schwarzschild, it led him to adopt a "pseudo-polar" coordinate system, making sure from the outset that this condition is fulfilled. He then expressed his result in terms of ordinary polar coordinates. The metric given by Schwarzschild is the unique spherically symmetric solution of the field equation in a vacuum, and describes the gravitational field of a point mass at the origin. As we know today, it also describes a black hole, but that insight would only arrive some fifty years in the future.

Yet it was clear even then that there was a problem with the Schwarzschild line element at a particular value for the radial coordinate r, namely when $r = 2m$, where m stands for the mass creating the gravitational field. This value is now called the Schwarzschild

radius. For greater values of the radial coordinate there was no problem, but it seemed that space, for the Schwarzschild radius, in some sense ends because of the apparently singular behavior of the metric at this radius. Two months later Schwarzschild published what is now called the interior solution, which assumes that matter is given by a fluid sphere of constant density. According to this solution, pressure becomes infinite before the radius of the sphere is equal to the so-called Schwarzschild limit. In spite of these indications, which pointed in the direction of new physical phenomena connected with the Schwarzschild solution, these hints were not pursued at the time. Instead, as observed by the historian Jean Eisenstaedt, who pioneered studies of the history of the Schwarzschild solution, a pragmatic attitude prevailed.[2] Schwarzschild's exterior solution was quickly adopted, also by Einstein himself, as the basis for calculations concerning two of the three principal tests of general relativity—the Mercury perihelion motion and the gravitational light deflection—and substituted Einstein's initial approximative method.

Not long after Schwarzschild, Johannes Droste, a student of Lorentz in Leyden, published a simpler derivation of the exterior solution, which he had derived prior to becoming familiar with Schwarzschild's work. Droste had earlier worked on the *Entwurf* theory and could now apply his knowledge to the final version of Einstein's theory. He dropped the determinant condition and re-derived Schwarzschild's solution not only in the now standard form, but also in other coordinate systems. He was entirely clear that the choice of coordinates was merely a matter of convenience but that one choice may be more appropriate than another. In a letter to his friend Paul Ehrenfest on 25 May 1917, Einstein wrote that he considered Droste's thesis, in which these results were first presented, to be "extraordinarily fine."[3]

Given the importance of the Schwarzschild solution for the astronomical predictions of general relativity, it quickly became a standard subject of the early investigations of the theory's consequences and also of the first textbooks. In 1917, David Hilbert and Hermann Weyl also published their own derivations of Schwarzschild's solution. Weyl introduced further mathematical simplifications and derived the metric from a variational principle. While these treatments revealed even more properties of the solution, the question of its singular features was not fully understood and indeed was actually blocked from sight with the rationale that they did not represent physical phenomena but rather limits of the theory.[4]

The first few years after the completion of general relativity saw a wealth of new results, also with regard to exact solutions. In connection with an investigation of the nature of the electron, Hans Reissner found the solution for the gravitational field of an electrically charged sphere, a result that Weyl also discovered and that was later generalized by the Finnish physicist Gunnar Nordström. Today this solution is called the Reissner-Nordström metric. In 1917, the Italian mathematician Tullio Levi-Civita investigated the gravitational field produced by a homogeneous electric or magnetic field, and he, and Weyl, derived solutions for symmetric charged and uncharged mass distributions. The relation between gravitation and electricity had been an old problem, as had been the question of whether gravitation might explain what holds the electron together. These investigations did not provide answers to such questions because, as Wolfgang Pauli summarized in his 1921 review article:

> the gravitational attraction exerted by an electron on another electron or on a charge element of its own is however always very much smaller than the electrostatic Coulomb repulsion . . . so that the electron can under no circumstances be held in equilibrium by the gravitational field . . . balancing its own repulsive forces.[5]

B. THE ENIGMATIC NATURE OF SINGULARITIES

Also in 1916, Ludwig Flamm managed to embed the spatial part of the Schwarzschild solution within a Euclidean space of four dimensions, a representation that suggests consideration only of masses whose radius is greater than the Schwarzschild radius. Given the astronomical knowledge of the time, this seemed the most reasonable approach. In his book *Space, Time, and Matter,* first published in 1918, Hermann Weyl wrote: "The gravitational radius is about 1.47 kilometers for the sun's mass and only 5 millimeters for the earth's."[6] And in his 1921 textbook, Max von Laue even claimed that every mass has a greater radius than the Schwarzschild radius.[7] Still, for years the singular character of the Schwarzschild radius remained a dogma for practically all of the relativist community. For David Hilbert, who in 1917 was the first to attempt a definition of the regularity of solutions, which later turned out to be deficient, the Schwarzschild solution displayed two singularities, one at the origin and another at the Schwarzschild radius. The latter claim was first challenged by Cornelius Lanczos in 1922:

> This example shows how little one can infer an actual singularity of the field from the singular behavior of the functions [of the metric], since it may be possible to remove the latter by a coordinate transformation.[8]

His paper, however, was ignored. Lanczos himself remained skeptical about the nature of singularities and was aware that this question had to be widely opened. Meanwhile, the question of what actually happens at the Schwarzschild radius remained a matter of speculation. In the same year, during a conference at the Collège de France in Paris, this question was raised by the French mathematician Jacques Hadamard. Einstein's reported response was that if that radius could actually be reached somewhere in the universe, "this would constitute a terrible disaster for the theory; and it would be very difficult *a priori* to say what would happen physically, because then the formula ceases to be applicable."[9] The next day Einstein returned with a calculation to argue that the singularity cannot be reached. His argument probably involved the above-mentioned Schwarzschild limit of the interior solution, which is larger than the Schwarzschild radius, so that one might claim that the latter cannot be reached on physical grounds, an argument going back to Schwarzschild himself. He ironically referred to the "Hadamard disaster." Much later, in 1939, Einstein devised another physical argument for why the Schwarzschild radius cannot be reached.[10]

In 1920, Arthur Eddington had also reached a similar conclusion by considering a thought experiment in which the Schwarzschild radius would be approached with a measuring rod: "There is a magic circle which no measurement can bring us inside. It is not unnatural that we should picture something obstructing our close approach."[11] In his famous 1923 textbook, *The Mathematical Theory of Relativity*, he derived a generalization of the Schwarzschild solution, including the cosmological constant, that also includes a solution that de Sitter had derived earlier. A year later, in 1924, Eddington derived what is now called the Schwarzschild solution in Finkelstein coordinates. This is a remarkable choice of coordinates because in this representation all gravitational potentials remain finite, but Eddington did not draw any further consequences concerning the alleged Schwarzschild singularity.

In a beautiful paper from 1933,[12] the Belgian physicist and priest Georges Lemaître derived another important solution of Einstein's equation: the spherically symmetric solution with null pressure (later known as the Lemaître-Tolman solution, and later still

as the Bondi solution). In a single system of coordinates, he appropriately described a model with a Friedman solution (see below) in the interior and a Schwarzschild solution at the exterior. Thus, he was the first to prove that the Schwarzschild solution actually remains regular even at the Schwarzschild radius:

> The singularity of the Schwarzschild field is thus a fictitious singularity similar to the one of the horizon of the center in the original form of universe of De Sitter.[13]

Einstein himself, however, continued to believe that the Schwarzschild radius somehow acted as a barrier. In 1939, he published a paper where he claimed for a star in formation, represented by a model of circulating particles, in which matter cannot be concentrated arbitrarily because otherwise the constituting particles would reach the velocity of light.[14] What he neglected was the possibility of a collapse starting not from such stable orbits of circulating particles but from an implosion of matter. During the formative years, the properties of the Schwarzschild solution were also probed along another line of investigation that today is considered a key tool for distinguishing the actual physical properties of such a solution from the possible artifacts due to the choice of particular coordinate systems: that is, by considering the behavior of trajectories of material particles and light rays in such a solution. This approach was indeed pursued by many of the previously mentioned protagonists, beginning with Schwarzschild and Droste via von Laue to lesser-known physicists working in the 1920s, such as Carlo de Jans. Although some of their results clearly hinted at the insight that at the Schwarzschild radius there was actually no impenetrable barrier preventing an object from falling into the interior, those hints were not pursued.[15] The overall perspective remained that of an unphysical, almost forbidden terrain that could not or needed not to be further explored. Remarkably, these seemingly marginal results connected much later with the exploration of trajectories provided an important starting point for the renewal of the consideration of the nature of the Schwarzschild radius and of singularities in general relativity. As we shall argue in the next section, the exploration of the nature of singularities was not only motivated by stellar collapse, however, but in particular also by cosmology.

C. EXACT SOLUTIONS AND COSMOLOGY

During the 1910s and 1920s, the major driving force behind the investigation of exact solutions was the question of the cosmological implications of general relativity. In chapter 5 below we shall discuss these implications in detail. Here, we just examine the important role of the early cosmological investigations for the understanding of exact solutions. This exploration began with the publication of Einstein's static cosmos in 1917 and his discussion with Willem de Sitter; it culminated in Lemaître's 1933 paper on the expanding universe quoted above. In the context of his discussion with Einstein on the possibility of a Machian interpretation of general relativity, de Sitter found a second cosmological solution without mass, a circumstance that made such an interpretation doubtful. Similar to the Schwarzschild solution, de Sitter's solution became an object of intense further investigation, both with regard to its possible astronomical implications and to its interpretation with the help of various coordinate representations. It suggested, in fact, observable redshifts of distant nebulae, which at that time had been measured by Vesto Slipher. Numerous researchers, such as Lanczos, Weyl, Richard C. Tolman, Howard P. Robertson, and Lemaître, explored the de Sitter solution, expressing it in coordinates

that revealed remarkable features; for instance, that it could be interpreted as a universe in which observers are expanding exponentially away from each other but in a static spacetime.

In 1922, the Russian mathematician Alexander Friedmann discovered expanding universe solutions with positive spatial curvature for a nonvanishing cosmological constant, albeit without deeply exploring their physical implications. Einstein's attitude to Friedmann's results was ambiguous. Two years later Friedmann also identified expanding universe solutions with hyperbolic sections. His expanding universe solutions with positive spatial curvature were rediscovered in 1927 by Lemaître, who now also worked out their physical consequences.

How did this investigation of the cosmological aspects of general relativity affect the search for exact solutions and their interpretation? In fact, the issue of exact solutions remained somewhat in the background of the cosmological discussions. As we shall see in more detail below, between 1917 and 1920, the main issue was whether Einstein's static cosmos (Universe A) or the de Sitter solution (Universe B) would be confirmed by astronomical observations. But as a matter of fact, the pursuit of expanding universe solutions by outsiders such as Friedmann and Lemaître proved crucial not only for the eventual success of general relativity in explaining observations pointing to an expanding universe, but also with regard to certain key issues in the interpretation of exact solutions.

For Lemaître, this was a matter of principle. In his 1933 paper, mentioned above, he explicitly emphasized the importance of the theme of exact solutions:

> The theory [of general relativity] may be developed in two different ways: through the study of exact solutions of the equations of gravitation, using simplified models, or through approximations to the solution for more complicated problems. I think it is important not to mix up these two methods. In this paper, we will concern ourselves only with mathematically exact solutions.[16]

His approach was shaped by his comprehensive perspective, including the Schwarzschild solution, the intricacies of the interpretation of the de Sitter solution, and the cosmological solutions. The inclusion of the cosmological solutions in his reconsideration of the Schwarzschild solution made cosmology, as Jean Eisenstaedt put it, a "space for work, for thought"[17] that opened up insights not accessible to those who took a narrower viewpoint. In particular, Lemaître had apparently long been suspicious about the argument that the Schwarzschild limit, following from the interior Schwarzschild solution, implied the physical inaccessibility of the Schwarzschild radius in the exterior solution. After all, the interior Schwarzschild solution introduces specific assumptions about matter that are not without alternatives. His familiarity with the cosmological solutions made it clear to Lemaître that the Friedmann solution presents a possible alternative interior solution and suggested that there was actually no limit preventing such an interior solution from collapsing. Searching for a way to express the interior and the exterior solutions in a single coordinate system, Lemaître found an expression for the Schwarzschild line element that made it evident that there was no singularity at the Schwarzschild radius. Although widely ignored by the pioneers of general relativity, including Einstein, Lemaître's results became the seed of later work on collapsing stars, eventually leading to the concept of a black hole.

The ensuing low-water-mark period of general relativity, which was dominated by a neo-Newtonian outlook on general relativity, left little room for such a comprehensive perspective on exact solutions, but a renewed effort to look for such solutions and to

characterize their properties became a hallmark of the theory's renewal in the 1950s and 1960s. During this renaissance of general relativity, the physical interpretation of exact solutions became indeed a central concern of the emerging community of relativists, including studies of the nature of the initial singularity of Big Bang models and of the dynamics of collapsing stars. It was a renaissance also of the dispersed insights into the nature of exact solutions reached during the formative years. Only by the mid-1960s was it recognized that singularities are not flaws but a rather generic feature of solutions of general relativity with deep physical implications for its understanding.

NOTES

This chapter relies heavily on the work of Jean Eisenstaedt. We are grateful for his input and advice.

1. Einstein to Karl Schwarzschild, 9 January 1916, CPAE vol. 8, Doc. 181.
2. The following account is based on his studies, see Jean Eisenstaedt, "Histoire et singularités de la solution de Schwarzschild (1915–1923)," *Archive for History of Exact Sciences* 27 (1982): 157–198; "The Early Interpretation of the Schwarzschild Solution," in *Einstein and the History of General Relativity*, ed. D. Howard and J. Stachel, *Einstein Studies*, vol. 1 (Boston: Birkhäuser, 1989), pp. 213–233; and "Lemaître and the Schwarzschild Solution," in *The Attraction of Gravitation: New Studies in the History of General Relativity*, ed. J. Earman, M. Janssen and J. D. Norton, *Einstein Studies*, vol. 5 (Boston: Birkhäuser, 1993), pp. 353–389. See also Jean Eisenstaedt, "Trajectoires et impasses de la solution de Schwarzschild," *Archive for History of Exact Sciences* 37 (1987): 275–357; and Matthias Schemmel, "The Continuity between Classical and Relativistic Cosmology in the Work of Karl Schwarzschild," in *Genesis of General Relativity*, vol. 3, pp. 155–181.
3. Einstein to Paul Ehrenfest, 25 May 1917, CPAE vol. 8, Doc. 344.
4. See David Hilbert, "Die Grundlagen der Physik (Zweite Mitteilung)," *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse* 1 (1917): 53–76; and Hermann Weyl, "Zur Gravitationstheorie," *Annalen der Physik* 54 (1917): 117–145.
5. Wolfgang Pauli, *The Theory of Relativity* (New York: Dover, 1981), p. 171. Reprint of *The Theory of Relativity*, trans. G. Field (New York, Pergamon Press; Bombay: B. I. Publications, 1958). Originally published in German as "Relativitätstheorie," *Encyklopädie der matematischen Wissenschaften*, vol. V19 (Leipzig: Teubner, 1921).
6. Weyl, *Space, Time, Matter*.
7. Max von Laue, *Das Relativitätstheorie. Erster Band. Das Relativitätsprinzip der Lorentz transformation* (Braunschweig: Vieweg, 1921).
8. Cornelius Lanczos, "Ein vereinfachendes Koordinatensystem für die Einsteinschen Gravitationsgleichungen," *Physikalische Zeitschrift* 23 (1922): 537–539, p. 539. Translation from Eisenstaedt, "The Early Interpretation of the Schwarzschild Solution," p. 219.
9. CPAE, vol. 13, Doc. 131, p. 251, n. 11.
10. See John Earman and Jean Eisenstaedt, "Einstein and Singularities," *Studies in History and Philosophy of Science Part B* 30 (1999): 185–235.
11. Arthur S. Eddington, *Space, Time and Gravitation* (Cambridge: Cambridge University Press, 1920), p. 98. See also Eddington, *Report on the Relativity Theory of Gravitation*, Physical Society of London (London: Fleetway Press, 1918).
12. Georges Lemaître, "L'Univers en expansion," *Annales de la Société Scientifique de Bruxelles* A53 (1933): 51–85. For a historical discussion, see Eisenstaedt, "Lemaître and the Schwarzschild Solution."
13. Georges Lemaître, quoted in Odon Godart, "Contributions of Lemaître to General Relativity (1922–1934)," in *Studies in the History of General Relativity*, ed. J. Eisenstaedt and A. J. Kox, *Einstein Studies*, vol. 3 (Boston: Birkhäuser, 1992), pp. 437–452, p. 449.
14. Albert Einstein, "On a Stationary System with Spherical Symmetry Consisting of Many Gravitating Masses," *Annals of Mathematics,* 40:4 (October 1939): 936.
15. Eisenstaedt, "Trajectoires et impasses."
16. Lemaître, "L'Univers en expansion," p. 51.
17. Eisenstaedt, "Lemaître and the Schwarzschild Solution," p. 373.

4
EINSTEIN AND THE ASTRONOMERS

After completing general relativity, Einstein shared his joy and satisfaction with friends and colleagues. He was particularly pleased that the theory could, with great accuracy, account for the advance of Mercury's perihelion as observed in 1859 by the French astronomer Urbain Jean-Joseph Le Verrier (see box below). Einstein wrote to Arnold Sommerfeld:

> The result of the perihelion motion of Mercury gives me great satisfaction. How helpful to us here is astronomy's pedantic accuracy, which I often used to ridicule secretly![1]

Evidently, Einstein only gradually came to appreciate the merits of astronomy, just as it had taken him some time to realize the value of mathematics, which is evident from a letter he had written to Sommerfeld three years earlier when he needed help from his "mathematician friend" Marcel Grossmann:

> I am now working exclusively on the gravitation problem and believe that I can overcome all difficulties with the help of a mathematician friend of mine here. . . . I have gained enormous respect for mathematics, whose more subtle parts I considered until now, in my ignorance, as pure luxury![2]

MERCURY'S PERIHELION PRECESSION

Newton's theory of gravitation accounts for Kepler's observation that the planets move around the sun in elliptic orbits. The theory predicts that these orbits are fixed in space if one ignores the influence of the other planets. Astronomical observations indicated that the elliptic orbit of planet Mercury slowly rotates in space. Most of this rotation can be explained by the interference of the other planets and accounted for by Newtonian theory. But astronomers also discovered that there is a discrepancy between the observed value of this rotation and Newtonian theory by 43″ (arc seconds) per century.[3] This problem had already arisen in 1859, following the work of the French astronomer Urbain Le Verrier and remained unsolved until Einstein's general theory of relativity.

As early as December 1907, when Einstein was taking his initial steps on the way to a relativistic theory of gravitation, he realized that such a theory may provide an answer to this longstanding problem. To his friend Conrad Habicht he wrote:

> At the moment I am working on a relativistic analysis of the law of gravitation by means of which I hope to explain the still unexplained secular changes in the perihelion of Mercury.[4]

Six years later, shortly after the publication of the Einstein-Grossmann *Entwurf* theory, Einstein and his friend Michele Besso performed a calculation to test whether the new theory could account for the anomalous precession of the perihelion of Mercury and concluded that it could explain only 18″ of the discrepancy. At the time, this disappointing result did not cause Einstein to doubt the validity of the *Entwurf* theory. In fact, he never published this result and ignored it completely until November 1915, when he used it as one of the arguments to justify his decision to abandon the *Entwurf* theory.[5]

When in 1907 Einstein first conceived the idea of a relativistic theory of gravitation on the basis of the principle of equivalence, he realized not only that the theory might explain the rotation of Mercury's orbit but also that it implied a bending of light by gravitation and a gravitational redshift. Taken together, these effects are known today as the three classical tests of general relativity: the explanation of the perihelion shift of Mercury in 1915 was a first triumph of the theory; the confirmation of light bending in 1919 a second triumph; and the third, the test of the gravitational redshift, turned out to be a more intricate challenge.[6]

Einstein took a closer look at the possible astronomical consequences of his theory in the making only in 1911 when he held an academic position at the German University in Prague. Specific predictions of the theory that could be confirmed or rejected by direct observations were not only important in distinguishing it from alternative theories of gravitation but also particularly desirable to Einstein because of the daring and speculative character of his theory. He then published a paper, "On the Influence of Gravitation on the Propagation of Light,"[7] with which he hoped to draw the attention of astronomers to its implications.

In this paper Einstein considered light being bent by the gravitational field of the sun and predicted an angle of deflection of a light ray passing near the rim of the sun of 0.83″. The bending of light by gravity had been inferred from Newton's particle theory of light but fell into oblivion after the triumph of the wave theory of light. Einstein now re-derived it on the basis of the equivalence principle and Huygens's principle of wave optics. In the general theory of relativity, the concept of the bending of light is modified. The trajectory of a light ray is determined by a geodesic line in curved spacetime produced by the mass of the sun. The actual angle of deflection, 1.7″, is about twice the value predicted by Einstein in 1911.

A second prediction based on the equivalence principle is a change in the color of light due to the gravitational field, the so-called gravitational redshift. It is a consequence of the slowing of the rate of clocks in the vicinity of a massive body. The oscillations associated with light emitted from an atom may be viewed as a clock. The slowing of such a clock in a gravitational field means that the frequency of the light is reduced: the "color" of light is shifted to the red end of the spectrum. The predicted relative shift of the frequency of spectral lines was very small (2×10^{-6}), at the limit of precision of spectroscopic measurements at that time.

Einstein concluded his 1911 article stating:

> It is greatly to be desired that astronomers take up the questions broached here, even if the considerations here presented may appear insufficiently substantiated or even adventurous. Because apart from any theory, we must ask ourselves whether an influence of gravitational fields on the propagation of light can be detected with currently available instruments.[8]

To answer the last question, Einstein needed the input of astronomers and discussed the issue with Leo Wentzel Pollack, an astronomer working at the Institute for Cosmological Physics at the German University in Prague. Pollack wrote to Erwin Finlay Freundlich, who at that time was working as an assistant at the Royal Observatory (Königliche Sternwarte) in Berlin, about Einstein's two predictions, which "urgently need verification through astronomy,"[9] and that the only chance to observe these effects, if at all, was in the case of the sun. We do not have Freundlich's response but we know that it pleased Einstein. He wrote to Freundlich:

> Thank you so much for your letter, which naturally interested me very much. I would be delighted if you wished to tackle this interesting question.[10]

A few years later, in a letter to Karl Schwarzschild, Einstein praised Freundlich:

> He was the first astronomer to understand the significance of the general theory of relativity and to address enthusiastically the astronomical issues attached to it.[11]

Freundlich's interest in finding astronomical evidence for general relativity quickly brought him into conflict with his superiors and the more senior astronomers. In early 1915, for instance, he published a paper in which he criticized a hypothesis of the eminent Munich astronomer Hugo von Seeliger. Seeliger had claimed that ellipsoidal concentrations of small particles not only explained the zodiacal light but that one such concentration near the sun accounted for the perturbations of Mercury, which Einstein later explained on the basis of general relativity. When Seeliger in turn discovered a mistake in a paper by Freundlich on the gravitational redshift predicted by Einstein's theory, Seeliger angrily turned to Freundlich's superior, Georg Otto Hermann von Struve. Einstein attempted to help Freundlich, who wanted to be absolved from his normal duties to be able to spend all his time on general relativity.[12]

The long collaboration between Einstein and Freundlich suffered a setback caused by personal tensions and disagreements around the end of 1921. Nevertheless, they continued to collaborate until both of them had to leave Germany when the Nazis came to power. Freundlich, raised in the Protestant faith, had a Jewish grandmother, which was enough to decide his fate under the race laws of the new regime.[13]

A. THE BENDING OF LIGHT

Freundlich had the idea to look at photos of previous solar eclipses and to compare the positions of stars along the direction of their observation with and without the sun being

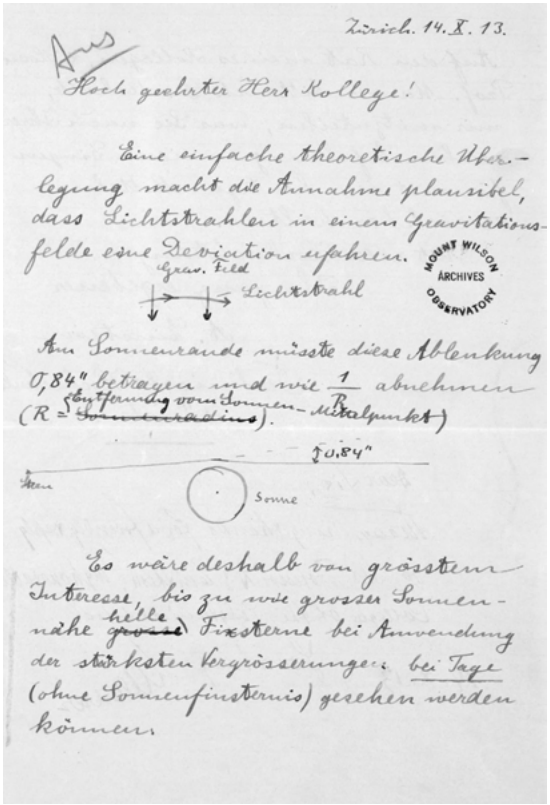

Zürich. 14. X. 13.
Hoch geehrter Herr Kollege!
Eine einfache theoretische Überlegung macht die Annahme plausibel, dass Lichtstrahlen in einem Gravitationsfelde eine Deviation erfahren.
Grav. Field
Lichtstrahl
MOUNT WILSON ARCHIVES OBSERVATORY
Am Sonnenrande müsste diese Ablenkung 0,84" betragen und wie 1/R abnehmen (R = Entfernung vom Sonnen-Mittelpunkt).
0.84"
Stern
Sonne
Es wäre deshalb von grösstem Interesse, bis zu wie grossen Sonnennähe helle Fixsterne bei Anwendung der stärksten Vergrösserungen bei Tage (ohne Sonnenfinsternis) gesehen werden können.

FIGURE 5: Einstein's letter to George Hale, 1913. By permission of the Huntington Library, San Marino, California.

present. A shift in their position would demonstrate the effect of the gravitational field of the sun on the propagation of the light rays from those stars, as predicted by Einstein. Freundlich soon lost hope that this would lead anywhere. Einstein thanked Freundlich for bringing the problem to the attention of astronomers:

> It is thanks to your zeal that the astronomers have now also started to show interest in the important question about the bending of light rays.[14]

In 1913, Einstein wrote to the astronomer George Hale, who at that time was director of the Mount Wilson Observatory, asking for his advice on the possibility of measuring the deflection of light in the vicinity of the solar rim (see figure 5). Hale's reply was that the only possibility to detect this effect would be during a solar eclipse. Einstein hoped that this could be done during the total solar eclipse, anticipated on 21 August 1914 in Crimea, Russia. Full of optimism, he wrote to Ernst Mach:

> Next year, during the solar eclipse, we shall learn whether light rays are deflected by the sun, or in other words, whether the underlying fundamental assumption of the equivalence of the acceleration of the reference system, on the one hand, and the gravitational field, on the other hand, is really correct.[15]

REVOLUTION IN SCIENCE.

NEW THEORY OF THE UNIVERSE.

NEWTONIAN IDEAS OVERTHROWN.

Yesterday afternoon in the rooms of the Royal Society, at a joint session of the Royal and Astronomical Societies, the results obtained by British observers of the total solar eclipse of May 29 were discussed.

FIGURE 6: The *London Times*, 7 November 1919.

In 1914, after the breakout of World War I, a German expedition was planned to observe this effect during a solar eclipse in the Ukraine, but it was interned for a brief period by the Russian authorities. As we have remarked earlier, at that time, the predicted angle of deflection was a factor of 2 smaller than the correct value of about 1.75″, which follows from general relativity and would be derived later.

The next serious effort to confirm Einstein's prediction was made during the total eclipse that occurred on 29 May 1919 along a narrow strip from South America over the Atlantic to Africa. The British Royal Astronomical Society organized two teams of observers and sent one, headed by Arthur Eddington, to Principe Island in the Gulf of Guinea, and the other one, headed by Charles Davidson, to Sobral in Brazil.[16]

Einstein's prediction was confirmed by astronomical observations during the solar eclipse of 1919, and he became a world celebrity overnight. It was not just the phenomenon itself, but also the measured angle that led to the sensational headline in the *London Times* on 7 November 1919: "Revolution in Science—New Theory of the Universe—Newtonian Ideas Overthrown" (figure 6).

Eddington wrote to Einstein that since the publication of the results, "all England has been talking about your theory. It has made a tremendous sensation."[17] Eddington, who advocated the improvement of scientific relations between formerly hostile powers during WWI, added:

> It is the best possible thing that could have happened for scientific relations between England and Germany.[18]

Einstein complimented Eddington on the achievement of his expedition:

> I congratulate you on the happy success of your difficult expedition. From the great interest you showed already in the past for relativity theory, I believe I may assume that it was due primarily to your initiative that the expeditions were carried out. I am astonished at the great interest the theory has generated among our English colleagues despite its complexity.[19]

A summary of the observations with a critical evaluation of the results of both teams was published in 1920 by Sir Frank W. Dyson, Astronomer Royal, Arthur Eddington, and Charles Davidson.[20] As final results, they gave the following deflections of starlight, as related to the radius of the Sun: Sobral 1.98 ± 0.12 arcseconds; Principe Island 1.61 ± 0.30 arcseconds. From this the authors conclude:

> Thus the results of the expedition to Sobral and Principe can leave little doubt that a deflection of light takes place in the neighborhood of the sun and that it is the amount demanded by Einstein's generalised theory of relativity.[21]

Summarizing this topic in *The Meaning of Relativity*, Einstein remarked that

> most careful preparations have been made to get more exact observational data at the solar eclipse in 1922. (p. 253 [93])

He referred to the solar eclipse that was expected to occur on 21 September 1922 and to be observed on a narrow belt of the earth from Somaliland across the Indian Ocean to Australia. After the initial excitement following the announced success of the British expedition, questions concerning the validity of the results were raised because, apparently, the mirror that reflected the light into the telescope camera was slightly bent. Therefore, the 1922 eclipse tests were seen as crucial in confirming or refuting relativity.

Observations were performed on Christmas Island in the Indian Ocean and at several sites in Australia. A team of Australian, Canadian, and American astronomers, led by Prof. William W. Campbell of the Lick Observatory in California, set up their instruments at Wallal, near Broome, on the northwest coast of Australia. Seven months later, on 12 April 1923, Einstein was informed by Campbell that the results from the Lick Expedition at Wallal had been the most successful. Overwhelming evidence was found that the light from 140 bright stars was bent at the expected angle as it passed near the sun (see figure 7).

B. GRAVITATIONAL LENSING

Considering the bending of light by a gravitational field, the idea that heavy bodies can act like lenses focusing the light of distant objects is not farfetched. Einstein explored this idea as early as April 1912, when he visited Berlin to discuss the astronomical consequences of his theory with Freundlich. Calculations in one of Einstein's notebooks dating from that period show that he explored this idea in some depth, deriving a lensing equation from which it follows that a terrestrial observer will see a multiple image of a lensed star, or, in the case of a perfect alignment, a so-called Einstein ring. Although Einstein did not publish these results at the time, he set great hopes in their potential role as an astronomical confirmation of general

OBTAIN FINE VIEWS OF SUN'S ECLIPSE

Astronomers in Australia See Corona 40,000 Miles Wide With Streamers 2,500,000 Long.

LATER THAN WAS SCHEDULED

Results Are Said to Favor the Einstein Theory, but Reports Will Be Long Delayed.

LONDON, Sept. 21 (Associated Press) —The astronomers had a splendid view of today's solar eclipse from Woolal, near Broome, Australia, according to a dispatch to the Times from Perth. The sky was cloudless and the observations were most successful.

The American expedition at Woolal, headed by Professor W. W. Campbell, director of Lick Observatory, saw a Corona 40,000 miles wide, from which four long streamers of light shot forth, one extending 2,500,000 miles from the sun's centre, says a dispatch to the Exchange Telegraph from Melbourne. The duration of the totality of the eclipse was four seconds shorter and began fifteen seconds later than had been predicted.

The American astronomers are said to be highly pleased at the success of the photographic plate, taken with the purpose of verifying Einstein's theory of relativity, namely, the bending path of a ray of light in its passage through the sun's gravitational field. They also expressed pleasure with the work of both the instruments and the observers.

The Canadian and Australian astronomers occupied stations near by the Americans. All had made careful efforts to verify the Einstein theory.

Christmas Island, where the British, Dutch and German expeditions were located, and Woolal, were considered the most favorable points or observation beceuse of the sun's high altitude there, but an expedition under George F. Dodwell, Government astronomer in South Australia, went to Cordillo Downs, in the Australian interior, whither it transported on the backs of camels more than 100 miles from the end of the railway line, a photographic telescope loaned by the Pittsburgh observatory. The observations made by this expedition also are said to have been most successful.

FIGURE 7: Clipping from *The Times*, 22 September 1922, reporting on the success of the solar eclipse observations.

relativity, even in the crucial period of its completion when his attention was focused on getting the correct field equation.

The meeting between Einstein and Freundlich took place at the time of the discovery of a nova star, the Nova Geminorum. It was discovered on 12 March 1912 by the astronomer Sigurd Enebo at Dombaas, Norway. In the following days, all major observatories as well as many amateur astronomers observed this new star. The observatory in Potsdam also took a number of photographs of this star between 15 March and 12 April. Freundlich, among others, was charged with photometric observations of this nova. As the historian of science Tilman Sauer has shown, the temporal proximity of the appearance of Nova Geminorum in 1912 with Einstein's Berlin visit during the week of 15–22 April suggests that this astronomical event was also discussed when he met Freundlich for the first time.[22]

It was also the observation of Nova Geminorum 1912 that must have motivated the depth with which Einstein explored gravitational lensing, including a calculation of the apparent magnification of the lensed star. He must have hoped to be able to explain in detail the light curve of the observed astronomical phenomenon. But as we know from his correspondence, he gave up this hope some time later. On 22 May 1916, he wrote to Emil Arnold Budde, a former director in the Berlin company Siemens & Halske, who, like so many others, had come up independently with the idea of gravitational lensing. He described how in the fall of 1915 the idea that the appearance of a nova could be explained by a distant star shining with an immensely increased intensity had even moved him to "joyous excitement" but, as he wrote at the end of the letter, that the joy had been "just as short as it had been great." Around this time, Einstein had once again met with Freundlich and must have discussed this idea further, eventually becoming more skeptical. On 15 October 1915, he wrote to his friend Heinrich Zangger:

> It has unfortunately become clear to me now that the "new stars" have nothing to do with the "lens effect," moreover that, taking into account the stellar densities existing in the sky, the latter must be such an incredibly rare phenomenon that it would probably be futile to expect one of the like.[23]

To Budde, he later gave the reasons why he had given up these ideas, pointing, in particular, to the behavior of the light curve during such an appearance.[24]

Einstein returned to the idea of gravitational lensing only twenty-four years later, in the context of a curious episode. The hero of this episode was the Czech amateur scientist Rudi W. Mandl, who in the spring of 1936 was looking desperately for somebody to consider his "proposed test for the relativity theory based on observations during eclipses of the stars."[25] He had been dismissed by several prominent scientists before he finally succeeded, on 17 April 1936, in meeting Einstein personally in Princeton. Mandl had developed a simple model describing how a star can focus the light of another star if both are aligned with the earth. On the basis of this model, he claimed to be able to explain not only the recently discovered annular shaped nebulae, but also the sudden extinction of dinosaurs, which he imagined to be caused by momentary intensification of cosmic radiation due to such stellar eclipses.

Einstein was as skeptical about such speculations as his colleagues who had earlier dismissed Mandl, but the core idea of gravitational lensing must have sounded familiar to his ears. Since Mandl was rather obstinate, Einstein eventually gave in and began to elaborate the idea. On 12 May 1936 he wrote to Mandl:

> Dear Mr. Mandl: I have calculated your intensification effect more precisely. The result is the following . . .[26]

Even on the issue of observability, Einstein was now slightly less pessimistic than he had been initially:

> In any case there may well be more chance to occasionally observe this intensification effect than the "halo effect" with which we have dealt earlier. But the probability that we can get so precisely into the connecting line of the centers of two stars at very different distances is rather low, even lower than the probability that the phenomenon, lasting in general only a few hours, happens to be observed.[27]

After much prodding by Mandl, Einstein finally agreed to publish a short paper on the matter. The note is entitled "Lens-Like Action of a Star by the Deviation of Light in the Gravitational Field" and was published in the 4 December issue of the journal *Science*.[28] It has since become not only the canonical starting point for the history of gravitational lensing, but has also had a lasting effect on subsequent research. Often, when a new phenomenon appeared in the sky, at least the possibility was considered that it might be related to gravitational lensing. Its observation, however, had to await the period of the renaissance of general relativity and the discovery of quasars, when the technical sophistication of astronomy and astrophysics had become fine enough to realize that gravitational lensing is an almost ubiquitous phenomenon and actually itself a powerful tool for exploring the universe.[29]

C. THE LONG STORY OF THE GRAVITATIONAL REDSHIFT

At the time of publication of his general theory of relativity, Einstein was convinced that the prediction of the gravitational redshift of spectral lines had already been confirmed by Freundlich. In October 1914, he wrote to his friend Paul Ehrenfest:

> Astronomer Freundlich has found a method to establish light refraction by Jupiter's gravitational field. In addition, he has established with astounding accuracy the shift of intensity centers of solar lines toward the red, based on an American's new, very exact observational data.[30]

In the same vein, he wrote to his former student Walter Dällenbach:

> One of the two important experimental consequences has incidentally already been verified splendidly, namely, the shift of spectral lines through a gravitational potential.[31]

Einstein shared with Sommerfeld his feeling that only personal intrigues prevented a conclusive verification of the predicted light bending effect, taking the confirmation of the redshift for granted:

> Freundlich has a method of measuring light deflection by Jupiter. Only the intrigues of pitiful persons prevent this last important test of the theory from being carried

> out. But this is not so painful to me anyway, because the theory seems to me to be adequately secure, especially also with regard to the qualitative verification of the spectral-line shift.[32]

Einstein was referring to the policy of Hermann Struve, the director of the Royal Prussian Observatory, who objected to Freundlich's work on this project.

Numerous measurements to test Einstein's theory were performed between 1915 and the early 1920s. In general, they did not confirm Einstein's claim. Specifically, the efforts of Charles St. John, a highly respected authority on the solar spectrum working at the Mount Wilson Observatory, produced consistently negative results. The failure to verify the existence of a gravitational redshift motivated some physicists to examine the implications of this result for the general theory of relativity. There were those who claimed that this was enough to falsify the whole theory; others argued that the general theory of relativity did not necessarily imply the gravitational redshift. The confusion generated by this situation is reflected in the following statement by St. John:

> The agreement of the observed advance of Mercury's perihelion and of the eclipse result of the British expeditions of 1919 with the deductions from the Einstein law of gravitation gives an increased importance to the observations on the displacements of the absorption lines in the solar spectrum relative to terrestrial sources, as the evidence on this deduction from the Einstein theory is at present contradictory. Particular interest, moreover, attaches to such observations, inasmuch as the mathematical physicists are not in agreement as to the validity of this deduction, and solar observations must eventually furnish the criterion.[33]

We shall not review here in any detail the convoluted history of the experimental and theoretical studies of this issue. For this, we refer the reader to an article by John Earman and Clark Glymour, "The Gravitational Red Shift as a Test of General Relativity: History and Analysis," and to the comprehensive book by the historian of science Klaus Hentschel, *The Einstein Tower*.[34]

Einstein was not deterred. He was convinced that general relativity necessarily implies the gravitational redshift and that, ultimately, it would be confirmed by experiments. He was encouraged by the result of Leonhard Grebe and Albert Bachem, two German physicists, who first (in 1919) obtained results that were about 80% of what was predicted by the theory. In further measurements, their results were even closer to those predicted by Einstein. To Eddington he wrote:

> Regarding the problem of the spectral redshift in the Sun, two young physicists, Grebe and Bachem in Bonn, have made a remarkable advance. . . . Through (objective) photometric analyses of the individual lines they found that most of the observed lines are asymmetric in the line-rich solar spectrum. Upon eliminating the distorted lines they find a quantity for the redshift that is in very good agreement with the theory.[35]

Eddington was not convinced. In reply to Einstein, he thanked him for the paper by Grebe and Bachem, which he found interesting and which looked convincing, but he suggested that he and Einstein should wait and see:

> I hear that St John has been making further researches, with magnesium and other lines, still getting a zero result; so I expect that for some time to come spectroscopists will be divided as to what the result really is.[36]

In an appendix that Einstein wrote in 1920 for the first English edition of *Relativity—The Special and the General Theory*, Einstein was more cautious when discussing the experimental efforts to confirm the predicted shifts in the frequency of spectral lines emitted by the sun. He reported there that "astronomers are working with great zeal towards the solution," but his conclusion was that the examination of available data did not allow a determination of whether observed redshifts of spectral lines could be referred to the effect of gravitation. Hoping that a definite conclusion would be reached within a few years, he stated:

> If the displacement of the spectral lines towards the red does not exist, then the general theory of relativity will be untenable.

In a later English edition of the same book, there is a footnote, added by the translator, saying that

> the displacement of spectral lines towards the red end of the spectrum was definitely established by Adams in 1924, by observations on the dense companion of Sirius, for which the effect is about thirty times greater than for the sun.

Walter Adams was the director of the Mount Wilson Observatory during Einstein's visits to Pasadena (see section D). It turned out that this conclusion also was premature.

Einstein also attached a vital importance to the confirmation of this effect in the Princeton lectures (see box below). At that time, he wrote to Michele Besso:

> The redshift of spectral lines is now beginning to be confirmed on all sides: . . . I didn't have a moment's doubt that it should be so; but others who believe they understand relativity theory certainly did.[37]

He was so convinced that he needed practically no experimental confirmation.

EINSTEIN SAYS THEORY IS LIABLE TO COLLAPSE

Dr. Einstein admitted yesterday that his theory of relativity would completely collapse if one of his conclusions can be disproved. The conclusion referred to was the one regarding the vibrations coming from the Sun through the chemical atoms filling space. "These vibrations," said Dr. Einstein, "would have their spectral lines bent slightly toward the red of the spectrum. If this is untrue, the whole theory is a fallacy."

DAILY PRINCETONIAN, 11 MAY 1921[38]

Einstein added an appendix to the second edition of *The Meaning of Relativity* in 1945. It was motivated by advances made in the theory of relativity, among which he mentions

> the conclusive demonstration of the existence of the red shift of spectral lines by the (negative) gravitational potential of the place of origin. This demonstration was made possible by the discovery of the so-called "white dwarfs" . . . this red shift was expected, by the theory, to be about 20 times as large as for the sun, and indeed it was demonstrated to be within the expected range.

The validity of this finding was criticized,[39] and the conclusive confirmation of the redshift result had to wait till the late 1950s, when R. V. Pound and G. A. Rebka compared the frequency of light emitted by a specific isotope of cobalt at the top and bottom of a 22.5-meter-high tower.[40] They measured a frequency shift that is equivalent to the slowing of a clock at the bottom of the tower by about one second in thirty million years. This experiment marked the beginning of an era when high-precision measurements were used to test the general theory of relativity. Nowadays, the effect of gravitation on the rate of clocks has to be taken into account in the timing system of Global Positioning System (GPS) technology.[41]

It is an irony of history that after many inconclusive attempts to verify Einstein's prediction of the gravitational redshift from astronomical observations, it was finally confirmed by a terrestrial experiment.

D. EINSTEIN VISITS CALTECH AND THE MOUNT WILSON OBSERVATORY

During the 1920s, Einstein corresponded extensively with de Sitter and Eddington and briefly interacted with Georges Lemaître. The latter three were well informed about contemporary work on observational astronomy, but Einstein himself was unaware of the progress made in observational methods and of the growing body of new results; he had had no contact with any of the scholars of that community. This situation changed during his first visit to Pasadena in January–February 1931, when he met a number of astronomers at Caltech and at the Mount Wilson Observatory with whom he discussed a broad range of astronomical observations related to the theory of general relativity and its cosmological implications. This visit was followed by two consecutive visits in January 1932 and in December 1932. The significance of these visits (particularly the first two) in the history of cosmology is discussed in chapter 5. We mention them here to complete the chapter on Einstein's interest in astronomy and on his interactions with astronomers during the formative years.[42]

Einstein went to Pasadena in the framework of an arrangement initiated by Robert Millikan, the president of Caltech between 1921 and 1945, according to which he would spend about two months of every winter term as a visiting professor at Caltech. Millikan, previously at the University of Chicago, experimentally confirmed Einstein's theory of the photoelectric effect based on the notion of light quanta. Although he doubted Einstein's theory of relativity, he had no doubts about Einstein's greatness as a scientist and wanted to bring him to Caltech even before this visit. It is possible that Einstein consented to visit this time in view of the deteriorating political situation in Germany, the rise of anti-Semitism, and the growing political strength of the Nazi party, which would make it impossible for him to remain in Germany, as was the case two years later. Einstein went to Pasadena accompanied by his wife, Elsa; his devoted secretary, Helen Dukas; and his scientific collaborator Walter Mayer.

As with his first visit to the United States in 1921, this one also generated great public interest and was covered extensively in the media. The *New York Times* assigned a journalist to report almost daily on all of Einstein's activities, statements, and lectures. On 3 January, they published an interview with Einstein and reported that he came to California with the expectation to get

> help from the scientists at Mount Wilson Observatory and the California Institute of Technology to solve the major problem of his mind, whether gravitation, light, electricity are not different forms of the same thing.[43]

The quest for a unified field theory was at that time a high priority on his scientific agenda, but the most significant impact of his visit was related to the fact that

> new observations by Hubble and Humason (astronomers at Mount Wilson) concerning the red shift of light in distant nebulae make presumptions near that the general structure of the universe is not static.
>
> Theoretical investigations made by La Maitre and Tolman (mathematicians of California Institute of Technology) show a view that fits well into general theory of relativity.[44]

"La Maitre" is a misspelling of Georges Lemaître, who was not at Caltech at that time but who met Einstein in 1927 and again during his second visit to Pasadena. Richard Tolman was the theoretical physicist with whom Einstein interacted during his visits in Pasadena and thereafter.

On 29 January, Einstein was treated to a visit to Mount Wilson to see the telescopes and meet with Hubble and other members of the observatory staff.[45] He first explored the telescope used for solar observations and then the 100-inch telescope, which was used for the ground-breaking observations by Hubble and Humason. This visit is commemorated in a historical photo of Einstein and others in front of the dome of the telescope (see figure 8).

A few days later, on 4 February, Einstein delivered a lecture at the offices of the Mount Wilson Observatory in Pasadena, not far from the Caltech campus. On the following day, this lecture made a front-page article in the *New York Times* and was also reported in the *Los Angeles Times*. The latter article appeared under sensational headlines: "Space Opened up by Einstein—Theory of Inclosed Creation May Be Modified—Scientists Speculate on Unified Field Equation—German Will Be Asked to Elaborate Views."[46] The article reports that when Einstein described the unified field theory, which combines electromagnetic and gravitational phenomena as the next step following the special and the general theory of relativity, he was asked to explain the relation between the unified field theory and cosmology.

> His answers, limited to a few sentences indicated, according to translators, that the first major result of the new theory will be a modification of his former idea that space is inclosed.

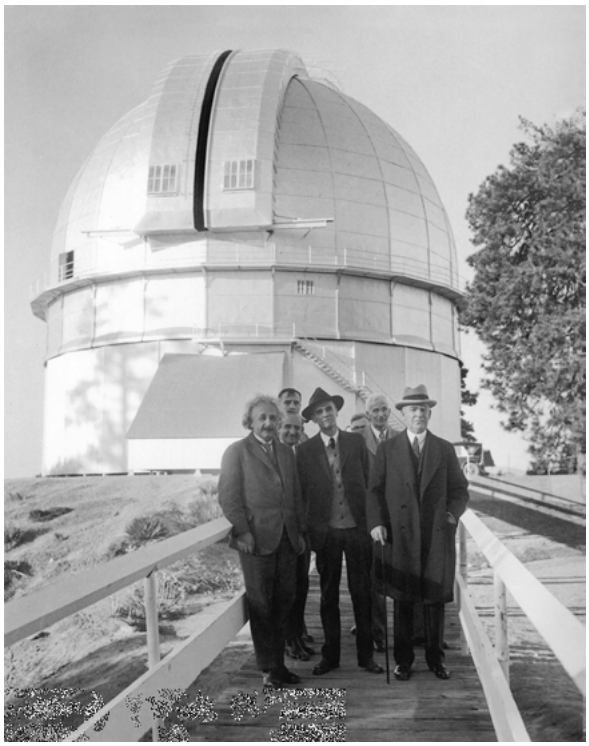


FIGURE 8: Albert Einstein at the 100″ Hooker telescope of Mount Wilson Observatory, Pasadena, with Walter S. Adams, William W. Campbell, Walther Mayer, and others, January 1931. Courtesy AP images.

However, several of the leading astronomers warned the readers not to draw premature conclusions:

> Dr. Walter S. Adams, Mt. Wilson Observatory director, and Dr. Charles E. St. John and Dr. Edwin Hubble, Einstein scientific collaborators, who were present voiced the opinion that the German theoretical physicist did not have time fully to explain his closing statement and cautioned the press against leaping to premature conclusions. Efforts will be made to arrange a second seminar to be devoted to Dr. Einstein's further elaborations of the possible revolutionary effect of his newest equation, recently announced here upon scientific conceptions of creation.

The second seminar took place on 11 February and the *New York Times*, on the following day, reported:

> Dr. Albert Einstein told astronomers and physicists here today that the secret of the universe was wrapped up in the red shift of distant nebulae. For more than an hour he discussed possibilities of the shape of the universe to take place of his discarded Einsteinian universe.[47]

The report quoted Einstein as saying: "The redshift of distant nebulae has smashed my old construction like a hammer blow." This was a blunt admission of Einstein's transition from

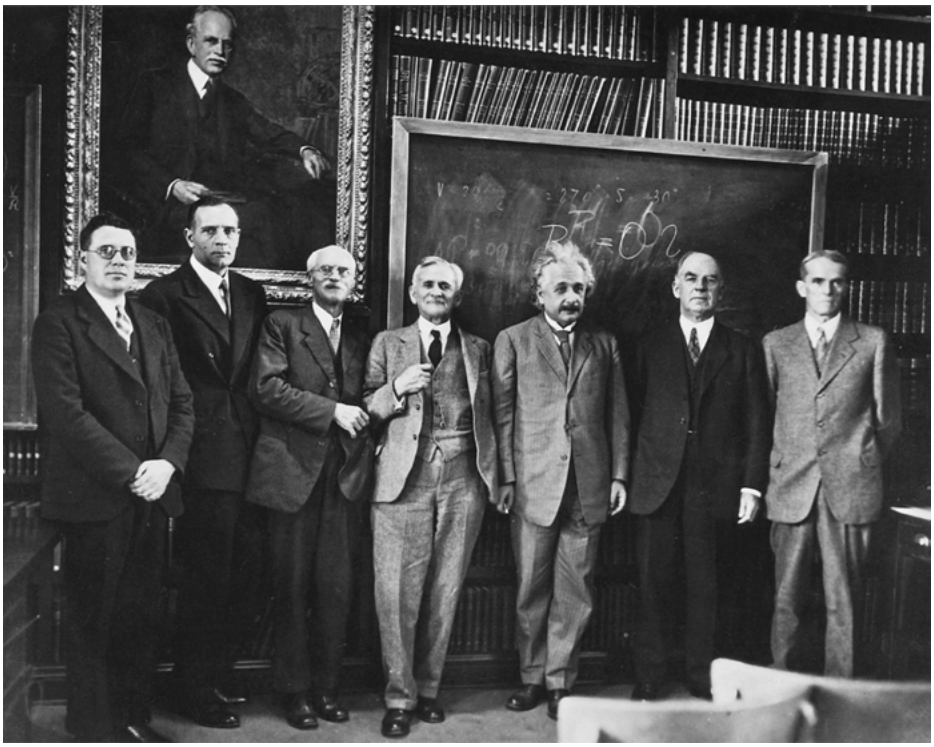


FIGURE 9: Milton L. Humason, Edwin P. Hubble, Charles E. St. John, Albert A. Michelson, Albert Einstein, William W. Campbell, and Walter S. Adams at Mount Wilson Observatory Library, January 1931. Courtesy Observatories of the Carnegie Institution for Science Collection at the Huntington Library, San Marino, CA.

his static closed model of the universe, which he had defended for more than a decade, to the expanding universe model.

In a statement summarizing Einstein's visit to Pasadena, Walter Adams emphasized the relation between general relativity and cosmology as one of the most important problems considered during that visit:

> In a clear and brilliant summary he (Einstein) has outlined the consequences of the different methods of treatment of the equations of generalized relativity and the aspects of cosmology to which they lead. The recent discoveries of the large red-shifts in the spectra of the most distant nebulae and the enormous apparent velocities which they indicate, together with the direct relationship found to exist between distance and amount of red-shift, have modified greatly some of the earlier views.[48]

We learn from this statement that not only questions of cosmology and of a unified field theory but also topics of ordinary solar astronomy were discussed during Einstein's visit:

> Of special interest to the members of the observatory staff have been . . . his suggestive discussion of the possible origin of circulatory movements in the solar atmosphere and their bearing on the rotation of the sun and the formation of sunspots.

Adams concluded his statement by idolizing Einstein's extraordinary intelligence and captivating personality:

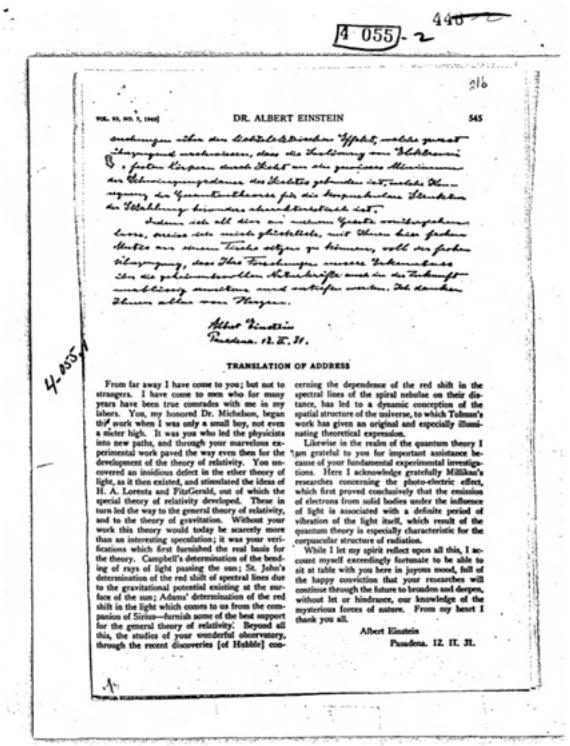
> Beyond any other consideration, however, has been the opportunity of observing his keen and brilliant mind at work, its instant grasp of the essentials of a problem and its extraordinary resourcefulness in its solution. Finally, no one can meet Prof. Einstein without an immediate realization of extraordinary intellectual power combined with a natural simplicity and kindliness which so often characterize great genius.

At the end of his visit, Einstein wrote a farewell address expressing gratitude for the two wonderful months he spent with excellent scientists and demonstrating his sense of humor:

> I saw worlds receding from us at unimaginable velocities, although their inhabitants certainly do not know us well enough so that such behavior could be justified.[49]

However, a more complete and eloquent summary of his visit is presented in an address at one of the receptions in his honor (see figure 10).

The second and third visits did not attract as much attention in the media as the first one. During the second visit, Einstein met with Willem de Sitter at Caltech and they published a joint paper. He also met with Georges Lemaître, who was then lecturing at Caltech. The third visit did not end as planned. In January 1933, Hitler rose to power



4 055 - 2

DR. ALBERT EINSTEIN 545

Albert Einstein
Pasadena. 12. II. 31.

TRANSLATION OF ADDRESS

From far away I have come to you; but not to strangers. I have come to men who for many years have been true comrades with me in my labors. You, my honored Dr. Michelson, began your work when I was only a small boy, not even a meter high. It was you who led the physicists into new paths, and through your marvelous experimental work paved the way even then for the development of the theory of relativity. You uncovered an insidious defect in the ether theory of light, as it then existed, and stimulated the ideas of H. A. Lorentz and FitzGerald, out of which the special theory of relativity developed. These in turn led the way to the general theory of relativity, and to the theory of gravitation. Without your work this theory would today be scarcely more than an interesting speculation; it was your verifications which first furnished the real basis for the theory. Campbell's determination of the bending of rays of light passing the sun; St. John's determination of the red shift of spectral lines due to the gravitational potential existing at the surface of the sun; Adams' determination of the red shift in the light which comes to us from the companion of Sirius—furnish some of the best support for the general theory of relativity. Beyond all this, the studies of your wonderful observatory, through the recent discoveries [of Hubble] concerning the dependence of the red shift in the spectral lines of the spiral nebulæ on their distance, has led to a dynamic conception of the spatial structure of the universe, to which Tolman's work has given an original and especially illuminating theoretical expression.

Likewise in the realm of the quantum theory I am grateful to you for important assistance because of your fundamental experimental investigations. Here I acknowledge gratefully Millikan's researches concerning the photo-electric effect, which first proved conclusively that the emission of electrons from solid bodies under the influence of light is associated with a definite period of vibration of the light itself, which result of the quantum theory is especially characteristic for the corpuscular structure of radiation.

While I let my spirit reflect upon all this, I account myself exceedingly fortunate to be able to sit at table with you here in joyous mood, full of the happy conviction that your researches will continue through the future to broaden and deepen, without let or hindrance, our knowledge of the mysterious forces of nature. From my heart I thank you all.

Albert Einstein
Pasadena. 12. II. 31.

FIGURE 10: Einstein's address at a reception held in his honor at Caltech, 1931 (AEA-4-055).

in Germany and Einstein could not go back to Berlin without endangering his personal safety. He stayed in Europe for several months until he got an entry permit for the United States. There he joined the newly established Institute for Advanced Study in Princeton, where he remained until the end of his life.

NOTES

1. Einstein to Arnold Sommerfeld, 9 December 1915, CPAE vol. 8, Doc. 161.
2. Einstein to Sommerfeld, 29 October 1912, CPAE vol. 5, Doc. 421.
3. For a historical account of the attempts by astronomers and physicists to explain this discrepancy, see Thomas Levenson, *The Hunt for Vulcan: . . . And How Albert Einstein Destroyed a Planet, Discovered Relativity, and Deciphered the Universe* (New York: Random House, 2015).
4. Einstein to Conrad Habicht, 24 December 1907, CPAE vol. 5, Doc. 69.
5. Meanwhile, the same result had been published by Johannes Droste in "On the Field of a Single Centre in Einstein's Theory of Gravitation," *Koninklijke Akademie van Wetenschappen te Amsterdam, Section of Sciences Proceedings* 17 (1914): 998–1011.
6. See Jeffrey Crelinsten, *Einstein's Jury: The Race to Test Relativity* (Princeton, NJ: Princeton University Press, 2006).
7. Reprinted in CPAE vol. 3, Doc. 23.
8. Ibid, p. 387.
9. Quoted in Klaus Hentschel, "Erwin Finlay Freundlich and Testing Einstein's Theory of Relativity," *Archive for History of Exact Sciences* 47 (1994): 143–201, p. 147.
10. Einstein to Erwin Freundlich, 1 September 1911, CPAE vol. 5, Doc. 281.
11. Einstein to Karl Schwarzschild, 9 January 1916, CPAE vol. 8, Doc. 181.
12. Hentschel, "Erwin Finlay Freundlich" p. 159.
13. For the scientific and personal relations between Einstein and Freundlich, see Hentschel, "Erwin Finlay Freundlich."
14. Einstein to Erwin Freundlich, mid-August 1913, CPAE vol. 5, Doc. 468.
15. Einstein to Ernst Mach, 25 June 1913, CPAE vol. 5, Doc. 448.
16. For the historical context of Eddington's expedition, see Matt Stanley, " 'An Expedition to Heal the Wounds of War': The 1919 Eclipse and Eddington as Quaker Adventurer," *Isis* 94 (2003): 57–89; and *Practical Mystic: Religion, Science, and A. S. Eddington* (Chicago: University of Chicago Press, 2007).
17. Arthur Eddington to Einstein, 1 December 1919, CPAE vol. 9, Doc. 186.
18. Ibid.
19. Einstein to Arthur Eddington, 15 December 1919, CPAE vol. 9, Doc. 216.
20. Frank Watson Dyson, Arthur Stanley Eddington, and Charles Davidson, "A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919," *Philosophical Transactions of the Royal Society of London A* 220 (1920): 291–333.
21. Ibid., p. 332.
22. Tilman Sauer, "Nova Geminorum 1912 and the Origin of the Idea of Gravitational Lensing," *Archive for History of Exact Sciences* 62 (2008): 1–22.
23. Einstein to Heinrich Zangger, 15 October 1915, CPAE, vol. 8, Doc. 130.
24. Sauer, "Nova Geminorum 1912."
25. Jürgen Renn and Tilmann Sauer, "Eclipses of the Stars—Mandl, Einstein, and the Early History of Gravitational Lensing," in *Revisiting the Foundations of Relativistic Physics—Festschrift in Honour of John Stachel*, ed. A. Ashtekar, R. Cohen, D. Howard, J. Renn, S. Sarkar, and A. Shimony (Dordrecht: Kluwer Academic Publishers, 2003), pp. 69–92, p. 69.
26. Einstein to Rudi Mandl, 12 May 1936, AEA 17034/35 (AEA: Albert Einstein Archives at the Hebrew University).
27. Ibid.
28. Albert Einstein, "Lens-Like Action of a Star by the Deviation of Light in the Gravitational Field," *Science* 84 (1936): 506–507. See also Jürgen Renn, Tilman Sauer, and John Stachel, "The Origin of Gravitational Lensing: A Postscript to Einstein's 1936 *Science* Paper," *Science* 275 (1997):184–186.

29. See Renn and Sauer, "Eclipses of the Stars."
30. Einstein to Paul Ehrenfest, before 10 April 1914, CPAE vol. 8, Doc. 2.
31. Einstein to Walter Dällenbach, 31 May 1915, CPAE vol. 8, Doc 87, p. 103.
32. Einstein to Arnold Sommerfeld, 28 November 1915, CPAE vol. 8, Doc. 153.
33. C. E. St. John, "The Displacement of the Solar Lines," *Nature* 106 (1921): 789–790.
34. John Earman and Clark Glymour, "The Gravitational Red Shift as a Test of General Relativity: History and Analysis," *Studies in History and Philosophy of Science* 11 (1980): 175–214; Klaus Hentschel, *The Einstein Tower: An Intertexture of Dynamic Construction, Relativity Theory, and Astronomy* (Stanford, CA: Stanford University Press, 1997).
35. Einstein to Arthur Eddington, 2 February 1920, CPAE vol. 9, Doc. 293.
36. Arthur Eddington to Einstein, 15 March 1920, CPAE vol. 9, Doc. 353.
37. Einstein to Michele Besso, before 30 May 1921, CPAE vol. 12, Doc. 141.
38. Quoted in Illy, *Albert Meets America*, p. 179.
39. See P.J.E. Peebles, "Robert Dicke and the Naissance of Experimental Gravity Physics," submitted to *European Physical Journal H*, March 2016. See also a talk given by Phillip J. E. Peebles, "The Revolution in Experimental Gravity Physics, 1957 to 1967," http://mediathek.mpiwg-berlin.mpg.de/mediathekPublic/2014-2015/Conferences-Workshops/A-Century-of-General-Relativity/Friday/Phillip-J--E--Peebles.html.
40. R. V. Pound and G. A. Rebka Jr., "Gravitational Red-Shift in Nuclear Resonance," *Physical Review Letters* 3 (1959): 439–441. The first reliable measurement of the gravitational redshift of a white dwarf was published in Daniel M. Popper, "Red Shift in the Spectrum of 40 Eridani B," *Astrophysical Journal* 120 (1954): 316.
41. See Clifford K. Will, *Was Einstein Right?* (Oxford: Oxford University Press, 1988). For a short explanation of the relation between relativity and GPS technology, see Clifford K. Will, *Einstein's Relativity and Everyday Life*, http://physicscentral.com/explore/writers/will.cfm.
42. For Einstein's visits to Pasadena, see David R. Topper, *How Einstein Created Relativity out of Physics and Astronomy* (New York: Springer Science+Business Media, 2013), chaps. 18, 23, 26; and Harry Nussbaumer, "Einstein's Conversion from his Static to an Expanding Universe," *European Physics Journal H.* 39 (2014): 37–62; Crelinsten, *Einstein's Jury*.
43. "Prof. Einstein Begins His Work at Mt. Wilson; Hoping to Solve Problems Touching Relativity," New York Times, 3 January 1921, p. 1.
44. Ibid.
45. For a description of this visit, see Marcia Bartusiak, *The Day We Found the Universe* (New York: Pantheon Books, 2009), chap. 16.
46. *Los Angeles Times*, 5 February 1931, p. A1.
47. "Red Shift of Nebulae a Puzzle, Says Einstein; He Tells Group of Astronomers That Bostonians' Discovery Upset One of His Theories," *New York Times*, 11 February 1921, p. 15.
48. AEA 90-387.
49. AEA 28-142.

5
THE GENESIS OF RELATIVISTIC COSMOLOGY

After completing his general theory of relativity, Einstein immediately realized its relevance to a description of the universe as a whole. Although none of the modern observations existed, the theory had some definitive implications for understanding the structure of the universe. Einstein's discussion of this subject, his debates with colleagues, and, particularly, his 1917 cosmology paper (see section B below) may be viewed as the genesis of modern cosmology. Until the creation of general relativity, the question of the geometrical properties of space at large and the question of how the structure of the universe is determined by gravity were separate issues. After Einstein's cosmology paper, they were closely interwoven.

General relativity is considered today as the theoretical basis of observational cosmology, but cosmology scarcely played a role in its genesis. Epistemological considerations and, in particular, Mach's criticism of Newton's absolute space were far more important to Einstein's thinking. As discussed in chapter 2, he followed Mach's idea of conceiving inertial effects as occurring in accelerated frames of reference not as a consequence of motion with respect to absolute space, but as being due to an interaction with distant masses. This idea played a key role in Einstein's heuristics. General relativity should explain all inertial effects by such interactions. Following this line of thought, Einstein's initial reflections on the "cosmological problem" are marked by his struggle to uphold Mach's approach.

A. THE EINSTEIN–DE SITTER DEBATE

In the fall of 1916, Einstein visited Leiden, where he met the Dutch astronomer Willem de Sitter, who in the midst of the war—which had also interrupted scientific communications—played a key role in making the general theory of relativity known outside Germany. De Sitter became Einstein's principal opponent in the discussion of the allegedly Machian features of general relativity. Since the solutions of the gravitational field equations are determined not only by the mass distribution represented by the right-hand side of the equation but also by boundary conditions, there was the

question of whether some trace of absolute space and time was still preserved in these boundary conditions. Einstein's first attempt to eliminate this problem was to assume that all the components of the metric tensor at infinity in the space dimensions are equal to 0 or ∞. He assumed that masses beyond the visible universe exist, which guarantee that these degenerate values of the field of inertia will evolve into the physical values at very large but finite distances.

The issue of boundary conditions sparked the Einstein–de Sitter debate, the beginning of which can be traced to a conversation between de Sitter, Einstein, Ehrenfest, and Nordström that took place in Leiden on 28 September 1916. De Sitter summarized Einstein's idea:

> Einstein wants the hypothesis of the boundedness of the world. What he means by that is that he makes the hypothesis (aware that this is an unprovable hypothesis) that at infinity (i.e., at very large, mathematically finite distance but farther away than any observable material body, hence outside the Milky Way etc.) there are masses—not observable though not in principle unobservable—such that the g_{ij} at infinity take on particular values (these do not have to be zero: that's impossible to say a priori), the same [values] in all coordinate systems. . . . He is therefore prepared to give up the complete freedom of transformations and, e.g., to restrict transformations to those in which there is always one time coordinate and three spatial coordinates, and in which a space axis never turns into a time axis etc.[1]

De Sitter then continued to speculate on the existence of such boundary conditions:

> Whether it is possible to find a set of degenerate values of g_{ij} which are invariant under a group of transformations that is not too restrictive is a question that can be settled mathematically. If the answer is no (as Ehrenfest and I suspect), then Einstein's hypothesis of boundedness is false. If the answer is yes, then the hypothesis is not in conflict with relativity theory. Still, I maintain that even in that case it is in conflict with the spirit of the relativity principle.

Einstein returned to Berlin, and with his assistant at that time, Jakob Grommer, he looked for a spherically symmetric solution with appropriate boundary conditions that, in accord with the principle of general relativity, would be independent of the coordinate system. In addition, they required that the resulting metric would have to be time independent and correspond to a static universe.

De Sitter strongly criticized this approach. In November 1916, he wrote to Einstein:

> I have been thinking much about the relativity of inertia and about the distant masses, and the longer I think about it, the more troubling your hypothesis becomes for me.[2]

Einstein responded a few days later:

> Now that the covariant field equations have been found, no motive remains to place such a great weight on the total relativity of inertia. . . . Which part of the inertia

> stems from the masses and which part from the boundary conditions depends on the choice of the boundary.[3]

It seems here Einstein has relaxed his categorical insistence on the Machian origin of inertia, but if so, it was not for long.

It took Einstein a few months to realize that his effort with Grommer was leading nowhere. He then found a clever way to solve the problem of boundary conditions and maintain a Machian basis of a physically acceptable cosmological model. He simply assumed that the universe is spatially closed and, hence, without boundaries. In February 1917, Einstein submitted for publication his first paper on the subject, his famous "Cosmological Considerations in the General Theory of Relativity," in which he stated:

> In a consistent theory of relativity there can be no inertia *relative to* "*space*," but only an inertia of masses *relative to one another*.[4]

About a week before submitting this paper for publication, Einstein wrote to de Sitter:

> Presently I am writing a paper on the boundary conditions in gravitation theory. I have completely abandoned my idea on the degeneration of the $g_{\mu\nu}$'s, which you rightly disputed. I am curious to see what you will say about the rather outlandish conception I have now set my sights on.[5]

To his friend Paul Ehrenfest in Leyden he wrote at about the same time:

> I have perpetuated something again as well in gravitation theory, which exposes me a bit to the danger of being committed to a madhouse. I hope there are none over there in Leyden, so that I can visit you again safely.[6]

B. EINSTEIN'S STATIC AND SPATIALLY CLOSED UNIVERSE

In his paper, Einstein discussed in detail his initial ideas on the boundary conditions to conclude why they had to be abandoned:

> For the system of the fixed stars no boundary conditions of the kind can come into question at all, as was also rightly emphasized by the astronomer de Sitter recently.[7]

He then introduced a spacetime model satisfying all his expectations concerning the constitution of the universe, including the explanation of its inertial properties by the distribution of masses acting as sources of the gravitational field. Another belief that guided Einstein in the construction of his first cosmological model was that the universe is static. (Actually, he may have first introduced the static nature of the universe merely as a simplifying assumption that only later, in the debate with De Sitter, acquired a greater relevance for his argument.) However, as it turned out, in the original field equations there are no spatially finite solutions representing a static homogeneous mass

distribution. The simple reason is that gravitational forces tend to contract the universe as a whole. He remarked:

> If it were certain that the field equations which I have hitherto employed were the only ones compatible with the postulate of general relativity, we should probably have to conclude that the theory of relativity does not admit the hypothesis of a spatially finite universe.[8]

These equations may be extended, however, by adding to the left-hand side of the field equation a term consisting of the metric tensor multiplied by a new constant, the *cosmological constant*. This term does not violate the general covariance, preserves the energy-momentum conservation, and essentially represents a small repulsive force acting on a large scale to balance the gravitational forces and make "possible a quasi-static distribution of matter."[9] Arthur Eddington later showed that the balance achieved in this way is unstable.[10]

Actually, Einstein could have had this additional term already in the gravitational field equation derived in "The Foundation of the General Theory of Relativity"[11] had he been looking for the most general form compatible with the four requirements: that the tensor on the left-hand side of the equation is second order; that it is formed from the metric components $g_{\mu\nu}$ and their first- and second-order derivatives and that it is linear in the latter. He must have been aware of this. We believe that, on aesthetic grounds, he would have discarded any term that was not absolutely necessary. He never liked this term, even when he was compelled to introduce it to allow for a static universe, because it complicated the theory and reduced its logical simplicity.[12]

To justify the introduction of the cosmological constant, Einstein shows at the beginning of his paper that also in Newton's theory of gravitation, an infinite static universe with a finite mass density is possible only by adding a similar constant on the left-hand side of the classical Poisson equation for the gravitational potential.

The presentation of these arguments in the paper probably follows the inverted order in which Einstein came to his conclusions and, in particular, to the introduction of the cosmological constant. In order to respond to de Sitter's objections against the introduction of invisible masses, Einstein must have first found the components of the metric describing his static and closed universe without boundaries. Since he then realized that it did not satisfy his original field equations, he tinkered with them, coming up with the idea to introduce the cosmological constant. Now he had to find good reasons for this audacious step in order to cope with the natural objection he could expect from de Sitter, namely, that this step was an entirely ad hoc measure serving only to defend his original conviction about the material origins of inertia. Finally, Einstein must have considered the analogue to his modification in Newton's theory and argued that a gravitational repulsion has to be included, both in Newtonian theory and in general relativity, to keep a static universe stable.

How do we know this? There is the legend that Einstein did some of his most important calculations on the back of an envelope. This is indeed a legend, as is made evident from the substantial notebooks that Einstein filled with lengthy calculations. But there is at least one case in which a calculation crucial for the early history of cosmology was made on the back of a letter, or rather, the draft of a letter (see box and figure 11).

LETTER TO THE AUSTRIAN EMPEROR KARL I

The letter was addressed to none other than the Austrian emperor Karl I. On 21 October 1916, the physicist Friedrich Adler, Einstein's colleague and neighbor in Zurich and secretary of the Austrian Social Democratic Party, had assassinated the Austrian prime minister, Count Karl von Stürgkh, as a protest against the war.[13] Adler was arrested and waited until May for his trial, at which he was sentenced to death, but he was eventually reprieved. Einstein wrote his draft, which began with the line "Your Majesty," decorated with playfully drawn ornaments. He continued:

> Under the pressure of an imperative obligation I am allowing myself the liberty, Your Majesty, to submit a request. . . . The political murder of which Fritz Adler made himself guilty has profoundly shaken the spirit of every right thinking person. I do not wish to varnish this horrible act as such with a single word. Considering the psychological state of the perpetrator, however, it seems that one deals rather with a tragic accident than with a crime.

After he described his personal relation to Adler, Einstein concluded:

> Under these circumstances I fulfill an imperative duty if I now ask you with all my heart, Your Majesty, to make use of the right to pardon, in case Adler is sentenced to death. A valuable life could thus be preserved.

Probably the letter was not sent because of the ongoing attempts to declare Adler insane, thus sparing him execution.

The back of the letter is filled with calculations in Einstein's hand, without any commentary (see figure 12). As in his published paper on cosmology, he started from a four-dimensional Euclidean space in which he defined a three-dimensional hypersurface representing a spherical space of constant curvature. By now considering the intrinsic geometry of this embedded hypersurface, he arrived at the line element (the metric) of his static universe. He next started to check whether this metric is actually a solution of his field equations, but the calculations break off. As Einstein then showed in his paper, it is not actually a solution but only suggests the introduction of the additional cosmological term. Although Einstein's paper bears clear hints of how he had originally proceeded to come up with this term, his "back-of-the-envelope calculation" confirms the remarkable inversion of the order of arguments by which, to use Reichenbach's language, the context of justification differs from the context of discovery.

Einstein's model describes a spherical, spatially closed, static universe with a uniform distribution of matter. The cosmological constant determines the average density of matter and the radius of the world. It has to be very small in order to maintain convergence to the Newtonian theory of gravitation in the limit of low velocities and weak gravitational fields. With this solution Einstein entirely avoided the problem of specifying appropriate boundary conditions, since a closed space does not have a boundary. He also believed that this model corresponded to a more or less realistic picture of the universe, as known

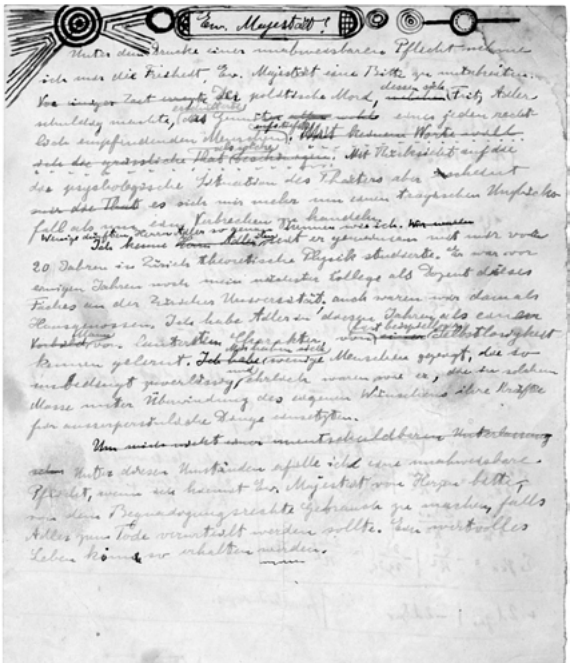


FIGURE 11: Draft of a letter to the Austrian emperor Karl I.

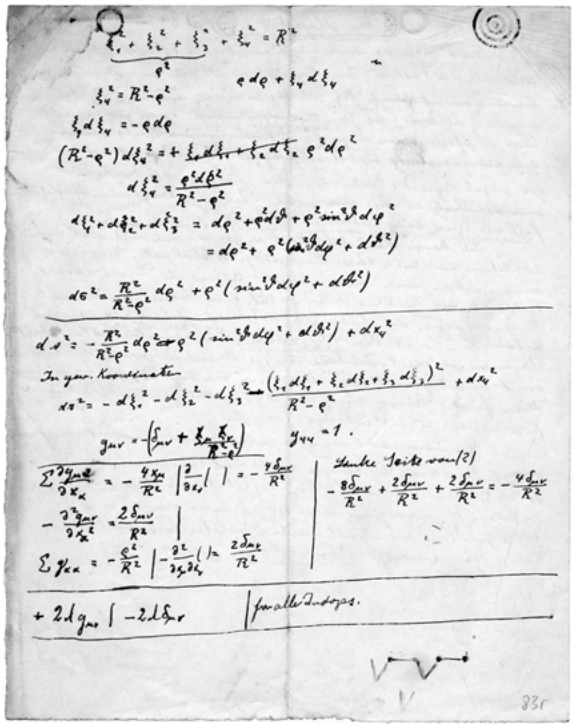


FIGURE 12: Einstein's "back-of-the-envelope calculation."

at that time. It was not even clear then that the galaxies observed outside our own Milky Way constituted objects of the same kind and that the universe actually extended far beyond it. However, Einstein described his model to de Sitter in an almost apologetic tone:

> From the standpoint of astronomy, of course, I have erected but a lofty castle in the air. For me, though, it was a burning question whether the relativity concept can be followed through to the finish or whether it leads to contradictions. I am satisfied now that I was able to think the idea through to completion without encountering contradictions. Now I am no longer plagued with the problem, while previously it gave me no peace.[14]

At about the same time, Einstein referred to his cosmological model in similar terms in a letter to Michele Besso:

> You will have received the "Cosmic Considerations." It is at the very least proof that general relativity can lead to a system free of contradictions. Before, there has always been the fear that the "infinite" harbored irreconcilable contradictions.[15]

The wording of these two letters supports the view that Einstein's paper on cosmology was motivated more by the epistemological debates about the general validity of the relativity of motion than by the desire to construct a cosmological model of the universe. The debate generated by this paper eventually led to the development of relativistic cosmology.[16]

De Sitter responded to Einstein:

> Well, if you do not want to impose your conception on reality, then we are in agreement. I have nothing against it as a contradiction-free chain of reasoning, and I even admire it. I cannot concur completely before having calculated with it, which is not possible for me to do right now.[17]

C. THE EINSTEIN–DE SITTER DEBATE CONTINUES

Shortly after Einstein's "Cosmological Considerations" was published, de Sitter demonstrated to Einstein that even the modified field equation allows a solution in which there is no matter acting as a source of the gravitational field. Nevertheless, test particles moving in this spacetime do have inertial properties that cannot be explained as an effect of Mach's distant stars.[18]

In the letter, de Sitter compared his cosmological model with that of Einstein by displaying the two solutions in two parallel columns. The heading of the column corresponding to Einstein's model reads: "*Three dimensional*, with supernatural masses"; the title of de Sitter's model is: "*Four dimensional*, without any masses" (see figure 13). In Einstein's model, only the spatial dimensions are confined to a sphere of radius R. In de Sitter's model, the entire spacetime is confined to a hypersphere of radius R. Three different mathematical expressions for the metric describing the two models are presented in the letter. The requirement that the metric is a solution of the gravitational field equation implies a relation between the cosmological constant, λ, and R, and a condition on the density of matter, ρ. In the published papers, de Sitter labeled Einstein's model as model A, and his own as model B. This is how they were referred to in the ensuing literature on this subject.

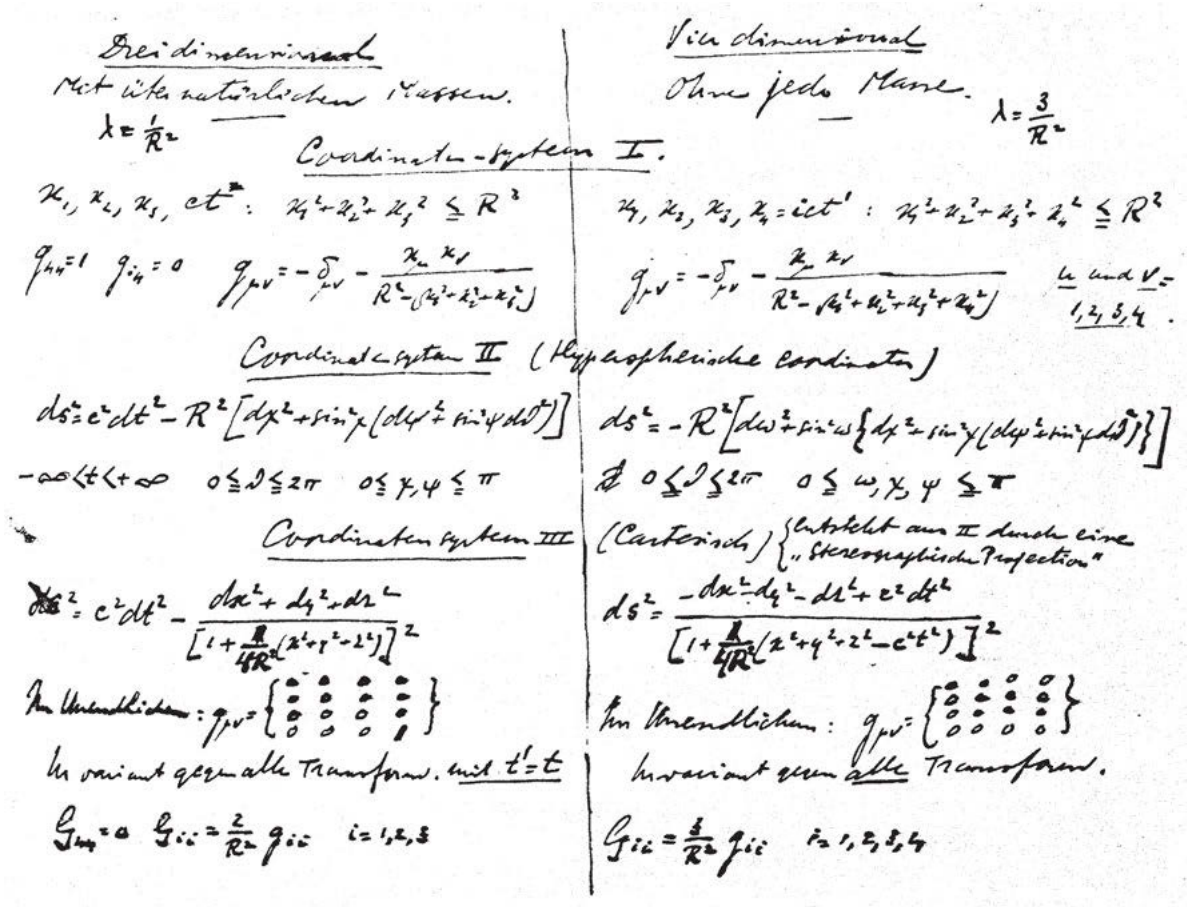


FIGURE 13: A comparison of the cosmological models of Einstein (left) and de Sitter (right).

Initially, model B was presented and perceived, just like model A, as corresponding to a static universe because the radius R is time independent. It turns out, however, that this is not the case. Model B behaves differently in different coordinate systems, and it is not easy to grasp the geometrical features, including apparent singularities, and their physical meaning in the different descriptions.

The correspondence and the published articles on de Sitter's model offer an example of the difficulties encountered in those days by scholars who tried to extract from the mathematical results of the theory their physical meaning. The mathematical tools, which would allow them to distinguish between features associated with a specific choice of coordinates and intrinsic features of spacetime, were developed only later. It was therefore not easy to interpret the physical meaning of particular solutions and to know if they contained genuine singularities.[19]

The correspondence between Einstein and de Sitter on model B, which lasted between November 1916 and April 1918, initially focused on the framework of a particular choice of coordinates that later turned out to be misleading. Initially they did discuss other

coordinate representations.[20] However, after a number of letters exchanged between them, in July 1916 they zeroed in on the static form of the de Sitter solution.[21] In this system of coordinates, time slows down as we move in the spatial direction, along the circumference of a circle of the radius of the universe, from an initial point O, so that it takes infinite time to reach the point at a quarter of the arc of the circle. Einstein could claim that all the mass is accumulated at this singular point (actually a line of points along the time coordinate) and that de Sitter's universe was not free of mass after all. On these grounds, Einstein rejected this solution as physically unacceptable in his "Critical Comment on a Solution of the Gravitational Field Equations Given by Mr. de Sitter," because it had a "genuine singularity . . . in the finite domain."[22] In a letter to de Sitter, in addition to claiming the existence of singularities, Einstein remarked:

> In my opinion, it would be unsatisfactory if a world without matter were possible. Rather the $g_{\mu\nu}$-field should be *fully determined by matter and not be able to exist without the matter*. This is the core of what I mean by the requirement of the relativity of inertia. One could just as well speak of the "matter conditioning geometry." To me, as long as this requirement had not been fulfilled, the goal of general relativity was not yet completely achieved. This only came about with the λ term.[23]

De Sitter responded to this with a reference to the limited empirical knowledge about the universe then available:

> I must emphatically contest your assumption that the world is mechanically quasi-stationary. We only have a snapshot of the world, and we cannot and must *not* conclude from the fact that we do not see any large changes on this photograph that everything will always remain as at that instant when the picture was taken.[24]

The exchange of letters between Einstein and de Sitter was cordial and marked by mutual respect. The following remark by de Sitter attributes their different attitudes to differences in their "scientific character": "The main point in our 'difference in creed' is that you have a specific belief and I am skeptic."[25] With this remark, he was responding to Einstein's letter talking about their difference in opinion on cosmology:

> We should not devote all that much time to our difference of opinion, which is only a difference in creed, so to speak, since it would not be fruitfully spent. We should see the possibilities without *wishing*.[26]

About a year later, and after correspondence with the mathematicians Hermann Weyl and Felix Klein on de Sitter's model, Einstein received a letter from Klein showing that the singularity in de Sitter's solution is an artifact of a special choice of coordinates. This correspondence is part of what has become known as the Einstein–de Sitter–Weyl–Klein debate.[27] Einstein responded to Klein:

> You are entirely right. De Sitter's world is, in and of itself, free of singularities and its space-time points are all equivalent. . . . My critical remark about de Sitter's solution needs correction; a singularity-free solution of the gravitational equation without matter does in fact exist.[28]

Having agreed that the de Sitter solution was singularity free, Einstein still objected to it on the grounds that it was not static:

> However, under no condition could this world come into consideration as a physical possibility. For in this world, time t cannot be defined in such a way that the three-dimensional slices t = const. . . . are equal to one another (metrically).

Several years later, Einstein still objected to the non-static nature of de Sitter's model. On 23 May 1923, he disagreed with Weyl, who then preferred de Sitter's model:

> As regards the cosmological problem, I am not of your opinion. According to De Sitter, two material points sufficiently distant from each other will accelerate as they move apart. . . . If there isn't any quasi-static universe after all, then get rid of the cosmological term.[29]

Both Einstein's (model A) and de Sitter's (model B) cosmological solutions became the subject of intense debate and constituted the principal alternatives that were considered. Einstein, however, did not acknowledge in public what he privately communicated to Klein: that he agreed that de Sitter's universe was free of singularities.

Einstein's views up to this point, expressed in his extensive correspondence, demonstrate how strongly he was committed to Mach's ideas. Having realized that even the modified gravitational field equation admits solutions without matter, his expectation of a Machian explanation of inertia in the theory of general relativity had changed from a requirement imposed on the theory itself to a criterion of acceptability to be applied to specific solutions of the theory.

D. "COSMOLOGICAL SPECULATIONS" IN THE PRINCETON LECTURES

Einstein's cosmological reflections in *The Meaning of Relativity* were still totally based on his Machian conviction that matter plays a primary role and the concept of space a derived one. Discussing general relativity in the third Princeton lecture, Einstein first drew attention to the fact that one needs a field-theoretical context to successfully address Mach's challenge:

> It is contrary to the mode of thinking in science to conceive of a thing (the space-time continuum) which acts itself, but which cannot be acted upon. This is the reason why E. Mach was led to make the attempt to eliminate space as an active cause in the system of mechanics." (p. 215–16 [55–56])

He also recapitulated Mach's program:

> The properties of the space-time continuum which determine inertia must be regarded as field properties of space, analogous to the electromagnetic field. The concepts of classical mechanics afford no way of expressing this. For this reason Mach's attempt at a solution failed for the time being." (p. 216 [56])

A significant part of the final chapter of *The Meaning of Relativity* is devoted to the cosmological problem. From Edwin Adams's summary, which was published in the *New York Times* on 13 May 1921, we learn that the last Princeton lecture, "Cosmological Speculations," began with a discussion of the inherent difficulty of a static infinite universe filled with matter at constant density in Newton's theory (see box).

Adams's report of Einstein's lecture:

If we imagine a sphere of radius very large compared to the mean distance between the stars, our first view is that as we increase the radius of the sphere more and more a definite density of matter in the universe is approached. The astronomer Seeliger first showed that such a view is definitely opposed to the Newtonian law of gravitation, for this view immediately leads to the result that the gravitational field would also increase beyond all limits as we go out towards infinity, and this would mean that the stellar velocities would necessarily increase beyond all limits.

Adams continues:

Thus on the basis of Newton's theory we should have to conclude that the mean density of matter is zero. This could only be attained by assuming that the universe is an island floating in infinite space free of matter. But this view is wholly unsatisfactory, and Seeliger attempted to reconcile an infinite universe with finite density by assuming that matter of negative density is present in the universe. This assumption involves departure from Newton's law of gravitation, but no other argument leads to a similar conclusion, and so this is not a satisfactory solution.

There is no trace of such a discussion in the printed version of the lectures. Instead, Einstein begins by exploring the consequences of a model of an infinite universe, Euclidean at infinity (boundary conditions). In such a case, Mach would have been "wholly wrong in his thought that inertia, as well as gravitation, depends upon a kind of mutual action between bodies" (p. 259 [99]).

The challenges raised by de Sitter are not mentioned explicitly in *The Meaning of Relativity*, but the whole discussion is a defense of Einstein's position vis-à-vis that of de Sitter. Einstein claims that the gravitational field equation indicates that "Mach was on the right road in his thought that inertia depends upon a mutual action of matter." He mentions three effects, which follow from Mach's ideas and are implied by the theory, although their magnitude is too small to be detected experimentally. The three effects are as follows:

- The inertia of a body must increase in the neighborhood of material bodies.
- Accelerated masses exert an accelerating force on a neighboring body.
- A rotating hollow body must generate inside of itself a "Coriolis force" and a radial centrifugal force. (p. 260 [100])

Einstein had already considered the case of a large rotating hollow sphere in 1913 in the context of his *Entwurf* theory. It can be found in the pages constituting the Einstein-Besso manuscript on the perihelion motion of Mercury and in Einstein's Vienna lecture of September 1913.[30] These calculations showed that the metric field near the center of the shell exerts forces similar to the Coriolis force and the centrifugal force occurring in a rotating frame of reference. Einstein was convinced that these forces were not just similar but identical, and he continued to believe this even though Besso had already warned him shortly after he submitted the text for his Vienna lecture that this is not the case.[31] Einstein conflated two very different situations: (1) the (non-flat) metric field near the center of a large rotating hollow sphere and (2) the (flat) Minkowski metric near the center of a large nonrotating hollow sphere described in rotating coordinates. Einstein used the presumed equivalence of these two situations to give a Machian explanation of Newton's rotating bucket experiment.[32] This conflation persisted after November 1915, when Einstein replaced the non–generally covariant field equations of the *Entwurf* theory by generally covariant field equations. This can be seen in his correspondence with Besso in 1916 and in his correspondence with Hans Thirring in 1917.[33] Like Besso before him, Thirring found that situations 1 and 2 above are different. By the time of the Princeton lecture, if not earlier, Einstein had come to recognize that the forces in situation 1 are much smaller than those in situation 2 and that his Machian explanation of Newton's bucket experiment thus failed.[34] Meanwhile, in a paper coauthored with Josef Lense, Thirring had found that the metric field outside a solid rotating sphere should give rise to tiny forces similar to the Coriolis force. This effect is now known as the Lense-Thirring effect (see box).[35]

THE LENSE-THIRRING EFFECT

The Lense-Thirring effect, also known as the frame-dragging effect, is the effect of a non-static, specifically rotating mass, on the gravitational field, namely, the spacetime surrounding it. In Newtonian mechanics, on the one hand, the gravitational field generated by a material body depends only on its mass, not on its motion. General relativity, on the other hand, predicts that a rotating object drags with it space and time. This is analogous to the magnetic field produced by an electrically charged rotating body and is therefore known as gravito-magnetism. Lense and Thirring, like Einstein and Besso before them, calculated this effect in the interior of a rotating mass shell. Einstein and Besso also calculated the effect of solar rotation on the motion of the nodes of the planets. The calculations by Einstein and Besso were performed five years earlier, still in the framework of the *Entwurf* theory. Lense and Thirring were the first to perform this calculation in the final theory. The correspondence between Einstein and Thirring in 1918 indicates that Einstein played a significant role in the emergence of what is known today as the Lense-Thirring effect. In their paper in 1918, Lense and Thirring predicted that the rotation of a massive object would cause a precession of the orbit of a nearby test particle, and they discussed the possibility of testing their prediction on

the orbit of the closest moon of Jupiter. The Lense-Thirring effect was more recently confirmed with great accuracy, exploring another phenomenon caused by the frame-dragging effect. In 2004, NASA launched the Gravity Probe B experiment, which was designed to measure precisely the direction of the axes of rotation of four gyroscopes, located in a satellite orbiting about 640 km above the Earth. The direction of these gyroscopes is effected by two causes: the spacetime curvature caused by the mass of the earth and the spacetime dragging caused by its rotation. Measurements of the angles of the axes of the four gyros over the course of a year conformed accurately with the predictions of general relativity.

Einstein concludes this lecture with three arguments supporting a spatially bounded, or closed, universe and rejecting a spatially infinite universe (pp. 267–268 [107–108]):

- The assumption of a closed universe is much simpler than assuming appropriate boundary conditions at infinity.
- Mach's idea that inertia depends upon the mutual action of masses is realized to a first approximation in the equations of the theory of relativity. This suggests a Machian character of general relativity because it would be unsatisfactory to assume that inertia is caused partly by mutual interaction of masses and partly by an independent property of space (boundary conditions). Mach's idea can be implemented only in a finite universe bounded in space.
- An infinite universe, according to Einstein, is possible only if the mean density of matter in the universe vanishes. Such an assumption is less likely than that of a finite mean density.

Einstein concisely summarized his views on the relation between Mach's principle, the cosmological constant, and the model of the universe in his Nobel Prize address:

> Mach's stipulation can be accounted for in the general theory of relativity by regarding the world in spatial terms as finite and self-contained. This hypothesis also makes it possible to assume the mean density of matter in the world as finite, whereas in a spatially infinite (quasi-Euclidian) world it should disappear. It cannot, however, be concealed that to satisfy Mach's postulate in the manner referred to a term with no experimental basis whatsoever must be introduced into the field equations, which term logically is in no way determined by the other terms in the equations. For this reason this solution of the "cosmological problem" will not be completely satisfactory for the time being.[36]

This was the last time that Einstein fully defended his Machian conviction. After 1921 he realized that it cannot be generally correct that inertial effects can be explained exclusively by the presence of matter in this theory. His attitude toward Mach's ideas changed correspondingly, and the interpretation of the theory of general relativity along the lines

of Mach's philosophical critique of classical mechanics ceased to play a significant role in Einstein's research. This shift of interest was caused not least by the reorientation of his research program in the direction of a unified field theory of gravitation and electromagnetism, which had started already in the early 1920s (see chapter 8).

Nevertheless, the question of Mach's principle remained open since it was now closely associated with Einstein's cosmological ideas, which largely coincided with the thinking of his contemporaries. In fact, as we have mentioned earlier, in the period between 1917 and 1930, the prevailing topic of debate was which model of the universe represents a better representation of reality.

E. FROM A STATIC TO AN EXPANDING UNIVERSE

At the time of emergence of the first cosmological models by Einstein and de Sitter, there was significant progress in extragalactic astronomy. In 1917, the American astronomer Vesto Slipher, working at the Lowell Observatory in Flagstaff, Arizona, reported measurements of spectral lines from twenty-five objects outside of the Milky Way galaxy. They were called nebulae at the time but were later identified as galaxies. Slipher discovered spectral shifts toward the red, which were identified as Doppler shifts, indicating that these nebulae are receding from the earth at velocities determined by the magnitude of these redshifts.

Unlike Einstein, de Sitter was not content with erecting "a lofty castle in the air." He was looking for connections between his model and astronomical observations. In particular, his model predicted a redshift in spectral lines of light from distant objects with the magnitude of the redshift being a function of the distance. In retrospect, Slipher's observations and de Sitter's theoretical prediction could have led to the notion of an expanding universe, but it was too early for such a far-reaching conclusion. First of all, de Sitter's model was then still perceived to be as static as Einstein's model. Moreover, the nature of the predicted redshift and its dependence on distance was unclear and a matter of controversy.[37] Even if the observed redshifts could be related to the Doppler effect, indicating the recession of the light-emitting objects, this would not necessary mean the expansion of the underlying spacetime. Such an understanding of this phenomenon emerged only gradually within the next few years.

Einstein was either unaware of, or did not relate to, the developments in observational astrophysics. His first confrontation with the idea of an expanding universe was the paper by Alexander Friedmann published in 1922. Friedmann showed that the metric of Einstein's model with a time dependent curvature (radius) of the universe $R(t)$ satisfies Einstein's modified gravitational field equation (with the cosmological constant Λ). In this case, the coupled equations for the components of the metric tensor reduce to two simple differential equations for $R(t)$. The solutions of these equations describe dynamical models corresponding to a static, expanding, or contracting universe, depending on the value of Λ, which can be positive, negative, or zero. Einstein initially dismissed Friedmann's solution as involving a calculation error. In his "Comment on A. Friedmann's Paper: 'On the Curvature of Space,'" he wrote:

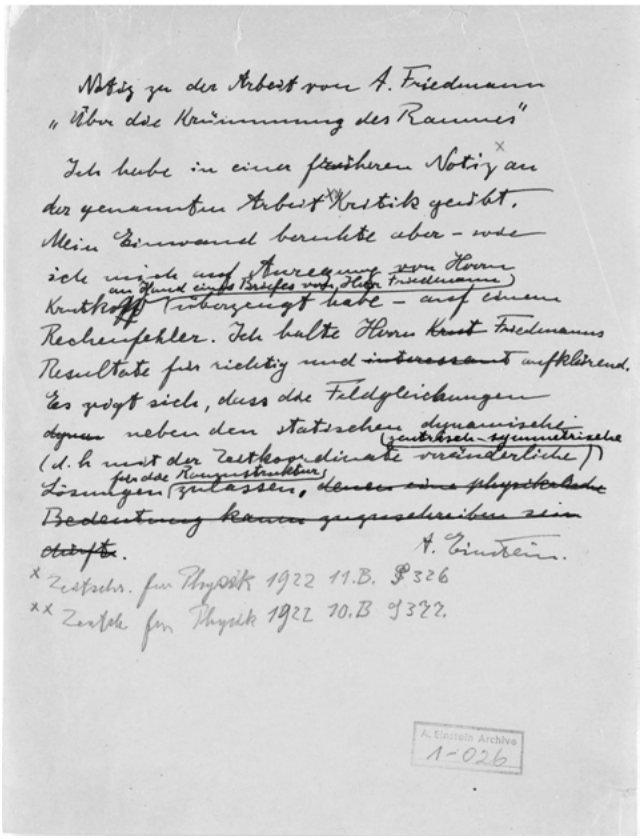
> The results obtained in the cited paper regarding a non-stationary universe seemed suspect to me. In fact, it turns out that the solution given does not agree with the field equations.[38]

When he recognized that this was not the case, he sent a note to the *Zeitschrift der Physik*:

> In an earlier note I exercised criticism on the mentioned paper [Friedmann's "On the Curvature of Space"]. My objection, however, was based on a calculational error—as I have become persuaded, at the suggestion of Mr. Krutkov, guided by a letter by Mr. Friedmann. I consider Mr. Friedmann's results correct and illuminating. It is demonstrated that the field equations permit, aside from the static solution, dynamic (i.e., variable with the time coordinate), centrally symmetrical solutions for the structure of space.[39]

In the handwritten version of this remark he added the words: "but a physical significance can hardly be ascribed to them," which he crossed out before sending it to the editor (see figure 14).

His reaction was equally dismissive when he learned about the rediscovery of an expanding universe solution by Georges Lemaître, who in 1927 published an article under the informative title "A Homogeneous Universe of Constant Mass and Growing Radius Accounting for the Radial Velocity of Extragalactic Nebulae."[40] This visionary paper remained unnoticed for several years, probably because it was published in a journal that


Notiz zu der Arbeit von A. Friedmann
„Über die Krümmung des Raumes"

Ich habe in einer früheren Notiz an der genannten Arbeit Kritik geübt. Mein Einwand beruhte aber – wie ich nicht auf Anregung von Herrn Krutkoff an Hand eines Briefes von Herrn Friedmann überzeugt habe – auf einem Rechenfehler. Ich halte Herrn Friedmanns Resultate für richtig und interessant aufklärend. Es zeigt sich, dass die Feldgleichungen neben den statischen dynamische (d.h. mit der Zeitkoordinate veränderliche) (zentrisch-symmetrische) Lösungen für die Raumstruktur zulassen, denen eine physikalische Bedeutung kaum zugeschrieben sein dürfte.

A. Einstein.

x Zeitschr. für Physik 1922 11.B. S 326
xx Zeitschr. für Physik 1922 10.B S 377.

A. Einstein Archive 1-026

FIGURE 14: Handwritten remark on Friedmann's 1922 work "On the Curvature of Space."

was not widely read. Two years before Hubble provided further observational evidence, Lemaître had already claimed that the galaxies are receding from the earth, not because they are moving through spacetime, but because spacetime itself is expanding at every point in the universe. Lemaître was also the first to provide an estimate of the rate of this expansion, which later became known as Hubble's constant.

Lemaître met Einstein in October 1927 at the Solvay Conference in Brussels. He later recalled that Einstein approved the mathematics but rejected the physical consequences: "Your calculations are correct but your physics is abominable."[41] Lemaître also had the impression that Einstein was not aware of astronomical facts and observations. Even when faced with Lemaître's thought-provoking arguments, Einstein could not imagine an expanding spacetime and continued to hold on to his perception of a static universe.

Einstein's change of attitude came with the accumulation of astronomical evidence in favor of an expanding universe, the decisive contribution being the work published in 1929 by Edwin Hubble, who was working at the Mount Wilson Observatory. A crucial element, which would allow cosmological conclusions to be derived from the redshift measurements, had been missing until then. There was no reliable estimate of the distance of the light-emitting objects. This missing element was provided in 1928 by Hubble, who developed a method to measure the distance to the nebulae and combine them with the redshift results. In 1929, he published the redshift and distance measurements of forty-six nebulae and showed that the further away they are, the faster they are receding from our galaxy. Hubble found a linear relation between the recession velocities and the distances. The ratio between the two is called Hubble's constant and is related to the age of the universe.

FIGURE 15: Einstein with the Belgian priest and astronomer Georges Lemaître, Pasadena, 1933. (akg-images, AKG124218)

Thanks to this result, Hubble has been credited with one of the most important discoveries in cosmology: the discovery of the expanding universe. This attribution is in fact to some extent an artifact of the way in which scientists tend to view and shape the historical record, as was shown by Helge Kragh and Robert W. Smith.[42] From a more nuanced historical perspective, the realization that we are living in an expanding universe was instead the result of the accumulation of a number of efforts, from the theoretical work of de Sitter, Friedmann, and Lemaître, via the pioneering astronomical observations by Slipher, to Hubble's establishment, jointly with his collaborator Milton Humason, of the distance relation.

Einstein learned about Hubble's results in early 1931 during a stay in Pasadena at Caltech. Although we have no direct evidence for that, it is reasonable to assume that his main discussion partner on the cosmological implications of Hubble's work was Richard Tolman. Tolman had joined Caltech in 1922, and by the time of Einstein's visit, his research interests had shifted from statistical mechanics and thermodynamics to general relativity and cosmology; he had become increasingly interested in the application of thermodynamics to relativistic systems and cosmology. Einstein and Tolman had much to discuss, and their communication was easy because Tolman spoke German. Hubble is not mentioned once in Einstein's diary from those days, but Tolman is. Their relationship evolved within a short time to a friendship that lasted well beyond their time at Caltech. When Einstein left Pasadena, he gave Tolman his photograph and a poem he wrote especially for him.

POEM FOR RICHARD TOLMAN, PASADENA 1930–1931

Manches uns verbinden taet	We share many an activity
Nicht nur Relativitaet	And not just relativity
So wenn der Skribenten Schar	So when the gaggle of greedy scribes
Gierig rings versammelt war	Were gathered with expectant vibes
Wie gings mir an solchem Ort	In such a place, how's my constitution
Ohne Ihr Erloesungsswort?	With ne'er a word of absolution?
Wenn durch hoehere Gewalten	If through uncontrollable forces
Du musst gar noch Reden halten	One is made to hold discourses
Deutsch bringt man es noch heraus	In German, you may still deliver
Aber Englisch—welch ein Graus!	But English—oh that makes one quiver!
Tolman machts und allemal	Tolman does, for once and all,
Besser als im Original!	It better than in the original!
Wenn ich mit so dummen Sachen	When I come with such foolish stuff
Zu ihm komme, muss ich lachen	To him, then I am forced to laugh
Freu' mich ueber sein Gesicht	His face I'm always pleased to see
Doch das sag' ich ihm dann nicht.	But that, of course, I keep for me.[43]

Toward the end of his visit, Einstein wrote to his friend Michele Besso:

> The people at Mount Wilson Observatory are outstanding. They have recently found that the spiral nebulae are approximately spatially uniformly distributed and show a strong Doppler effect proportional to their distance, which follows without constraints from the theory of relativity (without the cosmological constant). The problem is that the expansion of matter indicates the beginning at 10^{10} or 10^{11} years back in time. Since a different explanation of this effect poses great difficulties, the situation is very exciting.[44]

The "problem" to which Einstein referred is that the age of the universe estimated from these observations is less than the accepted age of stellar objects estimated by astrophysical methods. Based on Hubble's results, it was possible to estimate the age of the universe by calculating when all the observed objects started to recede from the same point. Hubble's initial result was 2×10^9 (two billion) years. This posed a serious problem because already in Hubble's time the age of the earth was estimated, based on radioactive evidence, to be between 1.6 and 3.4 billion years.[45] Einstein discusses this point in the appendix to the second edition of *The Meaning of Relativity* (p. 276 [116]). What is surprising here is that in the letter to Besso he erroneously specified the age of the universe as 10^{10} or 10^{11}. If this had been the result, there would have been no problem.

Almost immediately after his return to Berlin, Einstein published a short (less than three pages), hastily written paper on the cosmological problem in which he stated that Hubble's results had made his assumption of a static universe untenable.[46] This paper is a landmark in Einstein's cosmological considerations because it marks the culmination of the process of his conversion from his static to an expanding model of the universe.[47]

Einstein's main goal in this paper was to show that Hubble's results could be accounted for by the theory of general relativity:

> Now that it has become clear from Hubble's results that the extra-galactic nebulae are uniformly distributed throughout space and are in dilatory motion (at least if their systematic redshifts are to be interpreted as Doppler effects), assumption (2) concerning the static nature of space has no longer any justification and the question arises as to whether the general theory of relativity can account for these findings.[48]

Moreover, Einstein wanted to explore if this could be done without the cosmological constant, which to him was "theoretically unsatisfactory."[49]

To explore these questions, Einstein adopted Friedmann's model, which in 1922 he had branded as having no physical significance, and applied its analysis with the cosmological constant set to zero. Einstein assumed positive curvature and derived a differential equation for the radius of the universe. The solution of this equation led to a model of a universe that first expands with decreasing rate, reaches a maximum, and then contracts to zero radius. This result is equivalent to one period in Friedman's oscillating model, which Einstein adopted for a short time.

For Einstein the most significant feature of this result was the fact that it was obtained without the cosmological constant. His concluding remarks read:

> Most importantly, it seems that the general theory of relativity can account for Hubble's new facts more naturally (namely, without the λ term), than can the postulate of the quasi-static nature of space, which has now been rendered a remote possibility empirically.

This paper, at least temporarily, sealed the fate of the cosmological constant. A footnote in the appendix to the second edition of *The Meaning of Relativity* (this appendix will be discussed in section G) reads: "If Hubble's expansion had been discovered at the time of the creation of the general theory of relativity, the cosmologic member would never have been introduced" (p. 287 [127]).

One of the most popular myths about Einstein is that he referred to his introduction of the cosmological constant as the biggest mistake he had made in his life (see box). This footnote is probably the most explicit regret that Einstein expressed in writing.

EINSTEIN'S BIGGEST BLUNDER

The origin of the story about Einstein's biggest blunder can be traced to an article in *Scientific American* (1956) by George Gamow, in which the author recalls that, many years ago, he had heard Einstein's admission about this idea being "the biggest blunder he had made in his entire life."[50] Gamow repeated this claim in his autobiography, *My World Line*.[51] In his recent book, *Brilliant Blunders*, Mario Livio, an astrophysicist and author of popular science books, reports on the basis of an extensive investigation that there is no evidence that Einstein had actually made such a statement, either orally or in writing, and that it is most probably Gamow's invention. Yet this story has been quoted extensively; it appears in numerous books and articles and has become a generally accepted part of the Einstein scientific saga.

Einstein happily got rid of the cosmological constant, but not all his colleagues approved without reservation. In a letter to Einstein, Richard Tolman agreed that there are a number of arguments for setting the cosmological constant equal to zero, but that there is one strong argument against it:

> Since the introduction of the Λ-term provides the most general possible expression of the second order, which would have the right properties for the energy-momentum tensor, a definite assignment of the value $\Lambda = 0$, in the absence of experimental determination of its magnitude, seems arbitrary and not necessarily correct.[52]

Future developments showed that he was right. Today the Λ-term represents an important aspect of cosmology related to the mysterious dark energy.

The introduction of the cosmological constant should not be perceived as a mistake, because the additional term is really quite legitimate, and contrary to what Einstein wrote

in 1916, the resulting field equation was actually not the most general equation consistent with his demands. What might have been a mistake was to assign to it a specific value, with the intention to exclude the possibility of an expanding universe. Likewise, abandoning it, namely, fixing its value at zero, was probably an even more severe mistake. Today the cosmological constant plays an important role in explaining the accelerating expansion of the universe on the basis of the theory of general relativity.

F. FROM DISPUTE TO COOPERATION: THE EINSTEIN–DE SITTER UNIVERSE

Einstein's second visit to Pasadena at the beginning of 1932 overlapped with that of de Sitter (see figure 16). In Einstein's travel diary, we find on 8 January a brief entry:

> Drafted a note with De Sitter on the relation between the Hubble effect and the density of matter of a universe without curvature and without the Λ term (the cosmological constant).[53]

The "note," which they drafted, resulted in a model to become known as the Einstein–de Sitter universe. The published paper was even shorter than Einstein's paper discussed above. Its main features were the elimination of the cosmological constant and the adoption of a flat spatial geometry.[54]

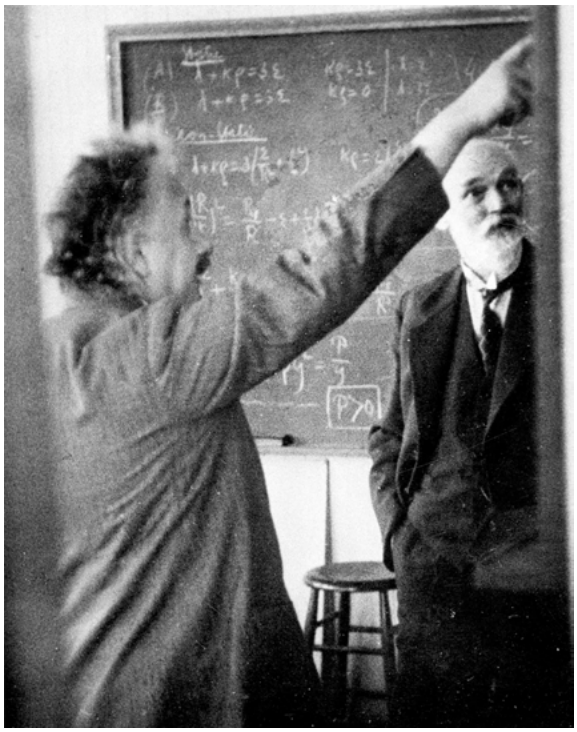


FIGURE 16: Albert Einstein and Willem de Sitter, Mount Wilson Observatory, January 1931. Copyright © AP Photos.

All the previous models of the universe had a finite spatial curvature. The exploration of a zero-curvature model was motivated by the German astronomer Otto Heckmann, who had pointed out that a finite density of matter in an expanding universe could also be compatible with a zero curvature of space. Einstein and de Sitter agreed that

> there is no direct observational evidence for the curvature, the only directly observed data being the mean density and the expansion . . . and the question arises whether it is possible to represent the observed facts without introducing a curvature at all.[55]

Indeed, the title of their paper, "On the Relation between the Expansion and the Mean Density of the Universe," indicates that this was their goal. Although the predicted density was higher than estimated by astronomical observations, the authors concluded that it was of the right order of magnitude and that "at the present time it is possible to represent the facts without assuming a curvature of three-dimensional space."[56]

The authors acknowledged that their assumption of zero curvature may change in the future: "The curvature is, however, essentially determinable, and an increase in the precision of the data derived from observations will enable us in the future to fix its sign and to determine its value."[57] They did not make a similar remark about their determination of the value of the cosmological constant to be zero.

The Einstein–de Sitter model, with zero space curvature, always expands, but the rate of expansion slows down to zero. In its initial version it described a matter-dominated universe. This aspect of the model had to be abandoned after the war, when research on microwave background radiation and on primordial nucleosynthesis provided evidence that the early universe had been permeated by radiation and only at a later stage evolved into a matter-dominated phase. This scenario is still compatible with the Einstein–de Sitter model if by the latter we mean a flat, spatially homogeneous, and isotropic universe with a zero cosmological constant. This version of the model became the most popular model in cosmology until almost the end of the twentieth century. It was abandoned when observational evidence about accelerated expansion compelled researchers to reintroduce the cosmological constant. The dramatic history of this controversy has been analyzed by the historian Helge Kragh.[58]

G. THE APPENDIX FOR THE SECOND EDITION OF THE MEANING OF RELATIVITY

When the second edition of *The Meaning of Relativity* was published in 1945, Einstein was fully, and almost exclusively, committed to his search for a unified field theory. In particular, since the early 1930s he had lost interest in cosmology and ignored the broad scope of theoretical and observational research in this field, which followed the work of Hubble. Still, he would not permit his views on the cosmological problem in 1921 to be his last word on the subject in 1945. He therefore added to this edition an appendix, "On the Cosmological Problem," which was also included in all subsequent editions. In this appendix, Einstein intended to discuss in detail the "so-called cosmologic problem":

> I feel urged toward a more exact discussion also by the fact that I cannot escape the impression that in the present treatment of this problem the most important basic points of view are not sufficiently stressed. (p. 270 [110])

Since Einstein's initial treatment of this problem in 1917, the basic question had been whether the general theory of relativity can account for a universe with an average homogenous density of matter. As we have seen, the exploration of this question had led Einstein in the past to introduce the cosmological constant to provide a negative pressure in order to counteract the gravitational contraction of the homogenous matter distribution. As was also remarked in the preceding sections, Einstein never liked this modification of the gravitational field equation because it "constitutes a complication of the theory, which seriously reduces its logical simplicity" (p. 271 [111]).

The main part of this appendix is an exposition of Friedmann's ideas, which showed that it is possible to have a finite density of matter in space without modifying the field equations. Moreover, his results were fully compatible with Hubble's observations leading to the discovery of expansion of the universe. The re-derivation of Friedmann's model is presented in this appendix in greater detail than in Einstein's 1931 paper, which is discussed in section E.

In an expanding universe model, extrapolating back in time one reaches a point at which the radius is zero and the density is infinite. Lemaître had concluded that this result indicates a physical beginning in his model, in the form of a primeval atom from which the universe evolved. This was the first version of the present-day notion of the Big Bang. Einstein was not bold enough to draw such a conclusion. For him the singular point at $G = 0$ (G – the radius of the universe) posed a problem:

> For every state of non-vanishing ("spatial") curvature, there exists . . . an initial state in which $G = 0$ where the expansion starts. Hence this is a section at which the density is infinite and the field is singular. The introduction of such a new singularity seems problematical in itself. (p. 284 [124])

Einstein hoped to remove this singularity by modeling the stars in the universe as particles in random motion, like molecules in a gas, hoping that such random motion would prevent the infinite condensation. Doing so, however, did not change the basic result.

As a final resort, Einstein blamed the existence of this singularity on the incompleteness of the theory. This view is succinctly expressed in the footnote:

> The present relativistic theory of gravitation is based on a separation of the concepts of "gravitational field" and of "matter." It may be plausible that the theory is for this reason inadequate for very high density of matter. It may well be the case that for a unified theory there would arise no singularity. (p. 284 [124])

This remark is in line with Einstein's preoccupation with the search for a unified field theory. He believed that singularities encountered in the general theory of relativity indicated its inadequacy and would be removed in a more complete theory.

Einstein concluded this appendix with nine remarks about Friedmann's universe and Hubble's findings, beginning with the "cosmological member," which "is to be rejected from the point of view of logical economy" (pp. 287–292 [127–132]). He discarded alternative explanations of the observed redshift of spectral lines from distant nebulae and asserted that the only viable explanation is "an expansion of the system of stars." This led Einstein

to the two main points, which have already been emphasized above. One is the singularity at the "beginning of the expansion," which he attributed again to the incompleteness of the theory "based on a division of the physical reality into a metric field (gravitation) on the one hand, and into an electromagnetic field and matter on the other hand." The other difficulty is the estimated age of the universe of about 10^9 years. This was incompatible with the contemporary astrophysical theories of the evolution of stars, which require longer times. This issue is mentioned three times in the concluding remarks. Einstein asserted that "the theory of evolution of the stars rests on weaker foundations than the field equations." However, a reliable estimate of the age of the earth crust also exceeds the estimated age of the universe. Einstein concluded his summary: "In this case I see no reasonable solution."

This contradiction between the age of the universe derived from Hubble's rate of expansion (Hubble's constant) and astrophysical estimates of the age of stars persisted until the late 1950s. It was then realized, mainly through the work of the astronomer Allan Sandage, that the distances to the galaxies were much longer than those found by Hubble, implying a slower rate of expansion. The present estimate of the age of the universe is 13.8 billion years.

* * *

Note that in this chapter our discussion extends beyond what we refer to as the formative years of general relativity and into the period known as "the low-water-mark period."[59] This characterization concerns both the epistemological and sociological status of the theory in this period. Concerning its epistemological status, work on the theory was largely characterized by a neo-Newtonian interpretation, which conceived relativity as providing small corrections to the Newtonian view, albeit with the notable exception of astronomical investigations into the issue of an expanding universe, a domain in which strong field effects were relevant. As far as the sociological aspect is concerned, many physicists working in this field shifted to other areas of research, specifically, to quantum mechanics. In this period, Einstein's theory generated more interest among mathematicians and astronomers than among physicists.

NOTES

1. We are grateful to Michel Janssen for drawing our attention to de Sitter's notes (in Dutch) of this conversation and allowing us to quote from his English translation of those notes. The material was made available to Janssen by Jan Guichelaar. See also Jan Guichelaar, *De Sitter: een alternatief voor Einsteins heelalmodel* (Utrecht: VBK Media, 2010).
2. Einstein to Willem de Sitter, 1 November 1916, CPAE vol. 8, Doc. 272.
3. Willem de Sitter to Einstein, 4 November 1916, CPAE vol. 8, Doc 273.
4. "Cosmological Considerations in the General Theory of Relativity," CPAE vol. 6, Doc. 43, p. 424.
5. Einstein to Willem de Sitter, 2 February 1917, CPAE vol. 8, Doc. 293.
6. Einstein to Paul Ehrenfest, 4 February 1917, CPAE vol. 8, Doc. 294.
7. "Cosmological Considerations," CPAE vol. 6, Doc. 43, p. 425.
8. Ibid., p. 430.
9. Ibid., p. 432.
10. Janssen, " 'No Success Like Failure,' " p. 201.
11. Albert Einstein, "Foundation of General Relativity," CPAE, vol. 6, Doc. 30, p. 180.

12. See section F.
13. See Peter Galison, "The Assassin of Relativity," in *Einstein for the 21st Century*: *His Legacy in Science, Art, and Modern Culture*, ed. P. Galison, G. Holton, and S. Schweber (Princeton, NJ: Princeton University Press, 2008), pp. 185–204.
14. Einstein to Willem de Sitter, 12 March 1917, CPAE, vol. 8, Doc. 311.
15. Einstein to Michele Besso, CPAE, vol. 8, Doc. 306.
16. Janssen, "'No Success Like Failure,'" section 5.
17. Willem de Sitter to Einstein, 15 March 1917, CPAE, vol. 8, Doc. 312.
18. Willem de Sitter to Einstein, 20 March 1917, CPAE vol. 8, Doc. 313.
19. Readers interested in the properties of de Sitter's model in different coordinate systems are referred to the articles by Michel Janssen and Christopher Smeenk in *The Cambridge Companion to Einstein*. See also David E. Rowe, "A Snapshot of Debates on Relativistic Cosmology, 1917–1924 (Part 1)," *Mathematical Intelligencer* 38 (June 2016): 46–58; and "A Snapshot of Debates on Relativistic Cosmology, 1917–1924 (Part 2)," *Mathematical Intelligencer* 38 (September 2016): 52–60.
20. See, in particular, CPAE, vol. 8, Docs. 317, 321, and 351.
21. Einstein to Willem de Sitter, CPAE vol. 8, Doc. 363. See "The Einstein-de Sitter-Weyl-Klein Debate," in CPAE vol. 8A (German version), Doc. 271, pp. 351–357. The above points can be found on pp. 353–354.
22. Albert Einstein, "Critical Comment on a Solution of the Gravitational Field Equations Given by Mr. de Sitter," CPAE, vol. 7, Doc. 5, p. 37.
23. Einstein to Willem de Sitter, 24 March 1917, CPAE vol. 8, Doc. 317, p. 309.
24. Willem de Sitter to Einstein, 1 April 1917, CPAE vol. 8, Doc. 321, p. 313.
25. Willem de Sitter to Einstein, 18 April 1917, CPAE vol. 8, Doc. 327.
26. Einstein to Willem de Sitter, 14 April 1917, CPAE vol. 8, Doc. 325.
27. "The Einstein–De Sitter–Weyl–Klein Debate," in CPAE vol. 8A (German version), pp. 351–357.
28. Einstein to Felix Klein, 20 June 1918, CPAE vol. 8, Doc. 567.
29. Einstein to Hermann Weyl, CPAE vol. 14, Doc. 40.
30. "Einstein and Besso: Manuscript on the Motion of the Perihelion of Mercury," CPAE vol. 4, Doc. 14, pp. 36–37; Einstein, "On the Present State of the Problem of Gravitation," CPAE vol. 4, Doc. 17, pp. 1260–1261).
31. See the so-called Besso memo, reproduced in facsimile in *The Genesis of General Relativity*, vol. 2, J. Norton, M. Janssen, J. Renn, T. Sauer, and J. Stachel (Dordrecht: Springer, 2007) pp. 786 and 789.
32. See Einstein's 1914 review article on the *Entwurf* theory, "The Formal Foundation of the General Theory of Relativity," CPAE vol. 6, Doc. 9, pp. 1031–1032.
33. Einstein to Besso, 31 July and 31 October 1916; Einstein to Thirring, 2 August and 17 December 1917, CPAE 8, Docs. 245, 270, 369, and 405.
34. For further discussion of Einstein's struggles with rotation, see Janssen, "'No Success Like Failure,'" sec. 4; and Janssen "The Twins and the Bucket," sec. 3.
35. Josef Lense and Hans Thirring, "Über den Einfluß der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie," *Physikalische Zeitschrift* 19 (1918): 156–163.
36. Albert Einstein, "Fundamental Ideas and Problems of the Theory of Relativity," in *Les Prix Nobel en 1921–1922*, ed. C. G. Santesson (Stockholm: Nobel Foundation, 1923), p. 489.
37. For a discussion of the redshift in de Sitter's model, see Christopher Smeenk, "Einstein's Role in the Creation of Relativistic Cosmology," in *The Cambridge Companion to Einstein*, pp. 228–269.
38. "Comment on A. Friedmann's Paper: 'On the Curvature of Space,'" CPAE vol. 13, Doc. 340.
39. "Note to the Paper by A. Friedmann 'On the Curvature of Space,'" CPAE vol. 14, Doc. 51.
40. Georges Lemaître, "Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extragalactiques," *Annales de la Société Scientifique de Bruxelles* 47A (1927): 49–59.
41. Lemaître's recollection of this discussion is reported without reference on p. 40 of Harry Nussbaumer, "Einstein's Conversion from his Static to an Expanding Universe," *European Physical Journal H* 39 (2014): 37–62.
42. Helge Kragh and Robert W. Smith, "Who Discovered the Expanding Universe?" *History of Science* 41 (2003): 141–162.
43. AEA 23-029. Translation by Lindy Divarci.

44. Einstein to Michele Besso, 1 March 1931, AEA 7-125.
45. G. Brent Dalrymple, "The Age of the Earth in the Twentieth Century: A Problem (Mostly) Solved," in *The Age of the Earth from 4004 BC to AD 2002*, ed. C.L.E. Lewis and S. J. Knell (London: The Geological Society, 1994), pp. 205–221, p. 207.
46. Albert Einstein, "Zum kosmologischen Problem der allgemeinen Relativitaetstheorie," *Sonderausgabe* (Berlin: Verlag der Akademie der Wissenschaften, 1931), pp. 235–237.
For an analysis of this paper in a broader context, with an English translation, see C. O'Raifeartaigh and B. McCann, "Einstein's Cosmic Model of 1931 Revisited: An Analysis and Translation of a Forgotten Model of the Universe," European Physical Journal H, 39 (2014): 63–85.
47. For a detailed description of this process, see Nussbaumer, "Einstein's Conversion."
48. English translation from C. O'Raifeartaigh and B. McCann, "Einstein's Cosmic Model of 1931 Revisited," p. 31.
49. Ibid., p. 6.
50. George Gamow, "The Evolutionary Universe," *Scientific American*, 1 September, 1956.
51. George Gamow, *My World Line: An Informal Autobiography* (New York: Viking Press, 1970). p. 44.
52. Richard Tolman to Einstein, 14 September 1931, AEA 23-031.
53. Albert Einstein, Travel Diaries–IV: USA, 3 December 1931 to 4 February 1932, AEA 29-136.
54. Albert Einstein and Willem de Sitter, "On the Relation between the Expansion and the Mean Density of the Universe," *Proceedings of the National Academy of Sciences* 18 (1932): 213–214.
55. Ibid., p. 213.
56. Ibid., p. 214.
57. Ibid., p. 214.
58. Helge S. Kragh, *Cosmology and Controversy: The Historical Development of Two Theories of the Universe* (Princeton, NJ: Princeton University Press, 1996).
59. See Jean Eisenstaedt, "La relativité générale à l'étiage: 1925–1955," *Archive for History of Exact Sciences* 35 (1986): 115–185; "Low Water Mark of General Relativity, 1925–1955," p. 277–292; and *The Curious History of Relativity: How Einstein's Theory of Gravity was Lost and Found Again* (Princeton, NJ: Princeton University Press, 2006).

6
THE CONTROVERSY OVER GRAVITATIONAL WAVES

With the completion of Einstein's formulation of the theory of general relativity, the action-at-a-distance interaction between material bodies in Newton's theory had been replaced by an interaction propagated by the gravitational field. Questions about the nature of this propagation and its velocity already naturally led to a discussion of gravitational waves. Such questions did in fact arise in connection with the *Entwurf* theory (see chapter 2, p. 42) after Einstein presented it in 1913 in Vienna.[1] In the discussion period Max Born asked:

> how fast does the effect of gravitation propagate according to your theory[?] It is not obvious to me that this would happen with the velocity of light; this must be a very complicated relationship.

to which Einstein responded:

> It is extraordinarily simple to write down the equation for the case where the disturbances one places into the field are infinitesimal. In that case the g differ only infinitesimally from those that would be present without that disturbance; disturbances propagate then with the same velocity as light.[2]

Yet when he presented his theory of general relativity on 25 November 1915 in Berlin, questions concerning the propagation of the gravitational field and the question of whether gravitational waves would constitute a consequence of his theory at first remained untouched. The first documented reaction of Einstein to this issue, with an explicit reference to such waves, appears in a short letter to Karl Schwarzschild, dated 19 February 1916, in which he laconically stated: "Thus there are no gravitational waves analogous to light waves."[3] He changed his mind about four months later, submitting on 22 June 1916 the first publication on gravitational waves within the general theory of relativity, "Approximative Integration of the Field Equations of Gravitation."[4]

A. THE EARLY INPUT OF MAX ABRAHAM

Before discussing the background of these developments and their impact on the thinking of Einstein and others on this topic in the midst of the debates marking the formative

years of general relativity, let us go back a few years to the role played by Max Abraham in the early stages of the search for a relativistic theory of gravitation. The papers he published between 1911 and 1915 contain some of the most detailed early descriptions of gravitational waves, comparing them with electromagnetic waves. Abraham was one of the great experts in the field of electrodynamics and electron theory. He is well known among historians of science due to the controversy he had with Einstein about what a field theory of gravitation should look like.

The work that Einstein did on gravitation during his stay in Prague between April 1911 and August 1912 was accomplished to a large extent in the context of his controversy with Max Abraham. In January 1912, Abraham was the first to publish a complete theory of the gravitational field, which was formulated within the framework of Minkowski's four-dimensional spacetime. At first, Einstein was impressed but later became skeptical. To his friend Besso he wrote: "At the first moment (for 14 days!) I too was completely "bluffed" by the beauty and simplicity of his formulas."[5] Although Abraham's approach turned out to be a dead end, he had developed some important insights from which Einstein also profited in the course of his own search for such a theory, possibly including thoughts about the nature of gravitational waves.

In a lecture presented in October 1912 and published the following year,[6] Abraham discussed the possibility of gravitational waves in a relativistic field theory of gravitation. He argued that the hope to detect gravitational waves was futile, basing his argument on a comparison with electromagnetic waves. The latter are produced by the accelerated motion of charges back and forth in a specific direction. Such a configuration of electric charges is known as an oscillating dipole moment. More complicated configurations of oscillating charges (like quadrupole and higher moments) also produce electromagnetic radiation, albeit at significantly lower intensities. Abraham argued that gravitational dipoles did not exist. He derived this conclusion from two fundamental principles of classical mechanics, the law of action and reaction and the equality of gravitational and inertial mass:

> Indeed, to impart an acceleration to one particle, another particle is necessary which, according to the law of action and reaction, is driven in the opposite direction. But now, the strength of the emitted gravitational wave depends on the sum of the products of the gravitational mass and the acceleration of the two particles, while, according to the reaction principle, the sum of the products of inertial mass and acceleration is equal to zero. Therefore, although the existence of gravitational waves is compatible with the assumed field mechanism, through the equality of gravitational and inertial mass the possibility of its production is practically excluded. It follows from this that the planetary system does not lose its mechanical energy through radiation, whereas an analogous system consisting of negative electrons circling around a positive nucleus gradually radiates its energy away. The life of the planetary system is thus not threatened by such a danger.[7]

Although Abraham did not consider the possibility of higher-order moments of gravitational waves, he was the first to achieve this fundamental insight into their nature.

B. THE CORRESPONDENCE WITH KARL SCHWARZSCHILD

Schwarzschild first contacted Einstein on 22 December 1915, sending him a letter from the eastern front during World War I.[8] In this letter he informed Einstein that he had completely solved the problem of the motion of Mercury's perihelion. As we discussed in chapter 3, this was the first exact solution of the field equations of general relativity for a single spherical nonrotating mass.

The next step in their correspondence is related to the existence of gravitational waves. From Einstein's letter mentioning gravitational waves, we conclude that it was a response to two letters from Schwarzschild. The first one, from 6 February 1916,[9] was about the properties of a special solution to the gravitational field equations, which later turned out to be crucial for the understanding of black holes. The other communication mentioned in this letter is lost, but we can be reasonably sure what it was about. Einstein showed little interest in the first letter, excusing himself for his belated answer because "the special cases discussed there awoke my interest to a lesser degree."[10] But he did react with excitement to something else that Schwarzschild had mentioned. It is now clear that a second communication from Schwarzschild had fired Einstein's curiosity anew: "But I find your new communication very interesting. I confirmed your calculation." This remark is followed by Einstein's first statement about gravitational waves quoted above. So what did Schwarzschild write that Einstein found so much more interesting than an exact solution to his field equations?

At about the same time, on 17 February 1916, as he was corresponding with Einstein, Schwarzschild also communicated with another colleague, Arnold Sommerfeld, a pioneer of quantum physics: "I am continuing to rummage around in Einstein's field equations. Today I am totally bewildered."[11] Schwarzschild used Einstein's theory to set up a differential equation for gravitational waves, from which he concluded: "Hence no wave motion but, infinite speed of propagation." Schwarzschild had taken the approximate field equation from the 4 November 1915 paper,[12] the first of the four papers that led to the final formulation of the theory. This approximation was obtained with a coordinate condition, which Einstein later dropped (see below). On this basis Schwarzschild tried to set up a wave equation. To his surprise he found an infinite speed of propagation, which of course constitutes a major problem and in any case excludes gravitational waves as he expected them. This must have been the same problem with which he also confronted Einstein in the letter now lost. Thus, the first to squarely address the issue of gravitational waves in general relativity was Schwarzschild.

Einstein repeated the calculation and he too was unable to find gravitational waves. In his response to Schwarzschild, Einstein blamed this failure on the approximation method, which he himself had introduced and later abandoned. He also referred to an argument first published by his one-time competitor Abraham against the very existence of such waves. The lack of gravitational waves "probably is also related to the one-sidedness of the sign of scalar T, incidentally. (Nonexistence of the "dipole.")"[13] though this was not his main argument. The "one-sidedness" of the sign of T (a parameter related to the energy-momentum of the field source) refers to the fact that there are no negative masses in the universe, while there are electrical charges of both signs.

Einstein had reached the conclusion that the approximation method used by Schwarzschild was not legitimate when he abandoned an earlier version of the theory

of gravitation that he had formulated on 4 November[14] and that formed the basis for Schwarzschild's argument. This approximation method involved a coordinate condition that was imposed by Einstein to restrict admissible coordinates in order to guarantee that the theory reduced to Newton's theory in the limit of weak and static gravitational fields. When Einstein derived the final version of the theory of relativity in late November 1915, he introduced another coordinate condition, $\sqrt{-g}=1$, this time in order to simplify the mathematical formalism. In the original handwritten version of the letter to Schwarzschild, Einstein added a footnote, emphasizing that the two coordinate conditions are not compatible, but Einstein now dropped the first one because he had realized that in the new version of the theory the Newtonian limit was attained differently from how he had previously thought. This is the meaning of his statement (in the same letter): "Since then, I have handled Newton's case differently, of course, according to the final theory." He erroneously believed that under these conditions there would be no gravitational waves.

C. THE CORRESPONDENCE WITH DE SITTER AND THE FIRST PAPER ON GRAVITATIONAL WAVES

On 22 June 1916, Einstein wrote to Willem de Sitter, his main interlocutor in the debate on the structure of the universe (chapter 5, p. 69 ff.):

> Your letter pleased me very much and inspired me tremendously. For I found that the gravitation equations in first-order approximation can be solved exactly by means of retarded potentials, if the condition of $\sqrt{-g}=1$ is abandoned. Your solution for the mass-point is then the result upon specialization to this case. Obviously your solution differs from my old one merely in the choice of coordinate system, but not intrinsically.[15]

The letter from de Sitter mentioned here has not survived, but we know what effect it had on Einstein. It spurred him to return to the challenge of gravitational waves. On the same day he wrote to de Sitter, he also submitted his milestone paper on the long road to understanding gravitational waves as part of our physical reality.[16] In the letter he summarized the main results of the paper.

Einstein's paper is based on an approximation scheme that is different from the one he used previously in his calculation of the precession of Mercury's perihelion. It assumes that the metric tensor describing the gravitational field deviates slightly from the tensor of flat spacetime (special relativity). The components of the tensor of deviations, $\gamma_{\mu\nu}$, are small quantities, so that any product of two of them can be neglected.

What was the role of de Sitter's letter and why could Einstein not use the same procedure beforehand? This is what must have happened: de Sitter sent him his solution for the mass-point at rest, which looked different from Einstein's own. The two solutions were intrinsically identical and differed only in the system of reference. This convinced Einstein of the legitimacy of dropping the $\sqrt{-g}=1$ condition, which was the main difference between his and de Sitter's reference systems. Earlier, because of this coordinate condition, Einstein thought he had to abandon his approximate linearized equations, which de Sitter's calculation had now shown him to be legitimate

after all. In the paper he made calculations that were very reminiscent of those he had performed previously, but he now basically ignored the issue of the Newtonian limit and the condition $\sqrt{-g}=1$.He found not only that he could get the same expression as de Sitter, but also that this solution corresponded to Einstein's own through a coordinate transformation. He could then address the failure of both his and Schwarzschild's method of deriving gravitational waves. All of this is summarized in an illuminating paragraph in Einstein's paper:

> We shall show that these $\gamma_{\mu\nu}$ [the small perturbative quantities] can be calculated in a manner analogous to that of retarded potentials in electrodynamics. From this follows next that gravitational fields propagate at the speed of light. . . . It turned out that my suggested choice of a system of reference . . . is not advantageous for the calculation of fields in first approximation. A note in a letter from the astronomer De Sitter alerted me to his finding that a choice of a reference system, different from the one I have previously given, leads to a simpler expression of the gravitational field of a mass point at rest.[17]

Einstein proceeded to transform the quantities $\gamma_{\mu\nu}$ into $\gamma'_{\mu\nu}$, which allowed him to write equations for the latter quantities, similar to the wave equations of electrodynamics. These equations could be solved exactly using the method of *retarded potentials* (see box), which is the familiar method used to derive electromagnetic waves. There are ten equations for the ten independent components of the gravitational field. These are connected by four relations, so that there are six solutions describing planar gravitational waves, namely, propagating fields with planar wave fronts. The six solutions can be divided into three types according to the symmetry of oscillations of the gravitational field with respect to the direction of motion. Two of them correspond to the familiar longitudinal and transversal modes of wave propagation. The third is a new type of symmetry, which was later referred to as *longitudinal-transversal*.

RETARDED POTENTIALS

The electromagnetic field at a point is generated by the charge and current values at all points in space, namely, by the charge and current distribution. If that distribution varies with time, then the field at a specific point is determined by the charges and currents at different points in space, at a later, or "retarded," time, which depends on the distance from that specific point. The retarded time is the time it takes for the field to propagate at the velocity of light, *c*, between the two points. Electromagnetic potentials involving retarded time are called retarded potentials and play an essential role in the solution of Maxwell's field equations and in electromagnetic radiation. The concept of retarded potentials applies also to the gravitational field in general relativity. Solutions of the linearized gravitational field equations are closely analogous to the electromagnetic retarded potentials.

Einstein then calculated the energy transported by the different wave types and concluded that the first two do not carry energy. There is a severe error in that derivation, which Einstein would correct only a year and a half later, but even then, the three types of waves would remain, two of them carrying zero energy. In the meantime, Einstein had to cope with

> the strange result that gravitational waves (types *a*, *b*, *c*), which transport no energy, could exist. The reason is that they are not "real" waves but rather "apparent" waves, initiated by the use of a system of reference whose origin of coordinates is subject to wavelike jitters.[18]

These "apparent" waves would not exist in a system of coordinates restricted by $\sqrt{-g}=1$. Thus, he concluded that the "unimodular" coordinate condition, which he abandoned to facilitate this calculation, "has a deep-seated physical meaning" after all.

Although the energy radiated in gravitational waves is extremely small, in his paper he commented:

> Nevertheless, due to the inneratomic movement of electrons, atoms would have to radiate not only electromagnetic but also gravitational energy, if only in tiny amounts. As this is hardly true in nature, it appears that quantum theory would have to modify not only MAXWELLIAN electrodynamics, but also the new theory of gravitation.[19]

The search for a quantum theory of gravitation has since then posed and continues to pose one of the major challenges at the frontier of theoretical physics.

D. CORRESPONDENCE WITH NORDSTRÖM AND THE SECOND PAPER ON GRAVITATIONAL WAVES

Following his correspondence with Gunnar Nordström, who was one the pioneers of relativistic gravitation theory, Einstein realized that there was something wrong with his 1916 paper. He had interacted with Nordström in the past and referred to his scalar relativistic theory of gravity as the only viable alternative to his own. In September 1917, Nordström sent a long letter to Einstein in which he described his problem with calculating the mass of a body in Einstein's general theory of relativity.[20] This is not a trivial question because a massive body generates a gravitational field and the energy of the field has its mass equivalent ($E = mc^2$), which is part of the mass of the body itself. Nordström found a way to calculate this total mass by integration over the space occupied by the body, which means that there is no energy in the gravitational field surrounding the body.

Nordström checked this result using the gravitational-field energy components derived in Einstein's paper and found, contrary to his expectation, that the field energy is nonzero. Einstein's response to this letter has not survived, but it prompted Nordström to rework his calculation based on the expressions for the gravitational field energy given in Einstein's seminal paper "The Foundation of the General Theory of Relativity." Nordström again found a discrepancy with the results based on Einstein's paper on gravitational waves. Nordström communicated his result to Einstein on 23 October 1917,

indicating that Einstein must have made a mistake: "The calculation is very simple, and one obtains . . . . With the best of intentions, I cannot get anything else, and I do not see where the inconsistency lies either."[21]

A short time after this letter, at the end of January 1918, Einstein submitted his follow-up paper, this time under the explicit title "On Gravitational Waves." The reason for this paper is stated at the beginning:

> The important question of how gravitational fields propagate was treated by me in an academy paper one and a half years ago. However, I have to return to the subject matter since my former presentation is not sufficiently transparent and, furthermore, is marred by a regrettable error in calculation.[22]

As the historian of science Michel Janssen has reconstructed, the "regrettable error" is a wrong derivation of the energy components of the gravitational field. In this calculation, Einstein used the auxiliary variables $\gamma'_{\mu\nu}$ instead of the correct components of the small deviations of the gravitational potential from their flat space values, $\gamma_{\mu\nu}$.

In the new paper, Einstein reworked the entire calculation in the framework of de Sitter's coordinates, correcting the "regrettable error" and its erroneous consequences. He then re-derived the three types of waves. Two of them carry no energy, as in the previous case, but now he was able to demonstrate that these waves can be obtained by a coordinate transformation from flat space. Thus, they are not real waves but mathematical artifacts of flat space described by curvilinear coordinates in spacetime:

> Those gravitational waves which transport no energy can, therefore, be generated from a field-free system by a mere coordinate transformation; their existence is (in this sense) only an *apparent one*."[23]

The field-free system corresponds to a flat spacetime in which there is no gravitational field and hence no gravitational waves.

One erroneous consequence of the "regrettable error" was that it allowed monopole and dipole radiation. This situation has also been rectified:

> This result shows that a mechanical system which permanently retains spherical symmetry cannot radiate; this is in contrast to the result of the previous paper, marred by an error in calculation.[24]

It is surprising that this result did not immediately raise Einstein's concern at the time. Had he remembered his interaction with Max Abraham, or had he applied a simple analogy with electromagnetism, he would have realized that there was something wrong with his derivation.

The most important result of this paper is the revised radiation formula, which describes the intensity and form of radiation emitted from a source of massive bodies. The gravitational waves are produced by an oscillating quadrupole moment of the mass configuration emitting the radiation (see box). A typical source of gravitational waves is a rotating binary system of celestial bodies. This formula is a landmark in the history of gravitational waves.

QUADRUPOLE RADIATION

Electromagnetic radiation is produced by positive and negative charges oscillating against each other along a single direction. This is the fundamental dipole radiation. More complicated oscillation schemes of charge configurations may give rise to higher order multipole radiation patterns, albeit of much weaker intensity. There is no analogue of gravitational dipole radiation. Gravitational waves are produced in leading order by a mass source changing along two perpendicular directions. This is what is known as quadrupole radiation. For example, a rotating mass distribution can be represented as combined oscillation along two perpendicular axes that are out of phase with each other. When the mass distribution expands along one axis, it contracts along the other. Thus, a common source of gravitational waves is a rotating binary system of stars or black holes.

E. THE DEBATE CONTINUES, BUT WITHOUT EINSTEIN

After submitting his 1918 paper on gravitational waves, Einstein remained silent on this issue for almost twenty years. Only in 1936 did he return to the issue of gravitational waves, in a joint paper with Nathan Rosen entitled "Do Gravitational Waves Exist?" They argued that in the exact theory there was no acceptable solution for gravitational fields. But as it turned out during the review process, their argument was flawed, and they had in fact found an exact solution for cylindrical gravitational waves.[25] But this is getting ahead of the story. During the formative years, Einstein reacted neither to further explorations of fundamental questions related to gravitational waves nor to criticism of his own work. Let us mention in this context the clarifying role of Arthur Eddington. We have already encountered him as the head of the British solar eclipse expedition that, in 1919, confirmed the bending of light by the gravitational field as predicted by Einstein, who thus gained celebrity status. In chapter 8, we will discuss Eddington's role in the search for a unified field theory. Less well known are Eddington's crucial contributions to the clarification of some of the intricate problems surrounding gravitational waves during the formative years. Indeed, this is yet another example of the recursive blindness of science, which tends to be oblivious to intermediate steps and to instead focus its gaze simply on the current state of events, at best connecting them to often-nebulous origins, in this case Einstein's 1916 seminal but problematic paper. That paper, but also the improved version of 1918, had actually left some fundamental questions open, such as what energy the gravitational waves actually transport and how to determine the gravitational waves emitted by more complex stellar configurations, such as a binary system. Likewise, the question of their speed of propagation had not been decisively settled either. Even after the intervention of Nordström, mentioned above, these questions remained open, and one of Eddington's primary achievements was to draw attention to such unsettled issues, which would continue to challenge the physics community for a long time to come.

Indeed, it is often not trivial to realize that an apparently evident claim actually poses a question. In 1916 Einstein had offered a solution to an approximate version of his field equations displaying gravitational waves propagating at the speed of light. In order to obtain this approximation, he had assumed a specific condition for the coordinates. He did not

explore whether the conclusion that gravitational waves propagate at the speed of light might possibly depend on that assumption, although this problem had already surfaced in Max Born's question in 1913, as well as in Schwarzschild's letter from 1916. Eddington now squarely addressed this issue. As he wrote in his first paper on gravitational waves:

> The problem of the propagation of disturbance of the gravitational field has been investigated by Einstein in 1916, and again in 1918. It has usually been inferred from his discussion that a change in the distribution of matter produces gravitational effects which are propagated with the speed of light; but I think Einstein really left the question of the speed of propagation rather indefinite. His analysis shows how the co-ordinates must be chosen if it is desired to represent the gravitational potentials as propagated with the speed of light; but there is nothing to indicate that the speed of light appears in the problem, except as the result of this arbitrary choice. So far as I know, the propagation of the absolute physical condition–the altered curvature of space-time–has not hitherto been discussed.[26]

It is indeed another characteristic of the formative years that the struggle with the meaning of coordinates continued (see chapter 1, p. 30; chapter 2, p. 43; chapter 7, p. 107). This struggle contributed at the same time to the rising significance of the geometrical interpretation of general relativity, which emphasized such notions as the curvature of spacetime mentioned by Eddington. Even the great Hermann Weyl, who had advanced some of the crucial concepts allowing such a geometrical interpretation, had missed this point when in 1919 he introduced in the third section of his famous textbook *Space, Time and Matter* a section on gravitational waves.[27] There he followed Einstein in discussing the three types of waves, but without noticing that two of them are actually spurious and only effects of the choice of coordinates.

Eddington, though, did not fail to explore precisely this point, noticing that the two unphysical types of waves "are merely sinuosities in the coordinate system, and the only speed of propagation relevant to them is 'the speed of thought.'"[28]

Eddington also re-derived Einstein's famous quadrupole formula, amending it by a factor of two and thus becoming the first author to publish the correct version and so consolidate Einstein's reasoning. In another paper, published in 1923 as well as in his textbook,[29] Eddington made further contributions to the investigation of gravitational waves and explored, in particular, the question of whether binary stars decay due to the emission of gravitational radiation, an issue that had already concerned Abraham and for which Eddington also had no conclusive answer. But in the history of science, the clear articulation of a question is often just as important as providing the answer, and this is surely what Eddington did when he stated: "The problem of two bodies in Einstein's theory remains an outstanding challenge to mathematicians like the problem of three bodies on Newton's theory."[30] This problem is still on the agenda of the relativistic theory of gravitation to the present day. For two black holes, a numerical solution was only achieved in 2005 after decades of research.

F. FROM THE FORMATIVE YEARS TO LIGO

After several years of intensive debates on the theory of general relativity and its consequences, interest in this field began to wane and receded into the background of research

in physics. This development marked the beginning of the low-water-mark period, which lasted from the 1920s until the 1950s.[31] The few physicists working on the various domains to which general relativity was relevant—cosmology, gravitational waves, mathematics, and experimental and observational investigations—were scattered throughout the world and only rarely exchanged results or ideas. There was, in other words, no recognized field of research called general relativity (and gravitation) as there is today. For a long time, the field had no continuous tradition and a correspondingly steady accumulation of knowledge.

This began to change only in the mid-1950s with the "renaissance of general relativity."[32] The beginning of the new period may be associated with the first international conference dedicated entirely to general relativity, which convened in 1955 in Bern. This was the fiftieth anniversary of special relativity and the year of Einstein's death. The Bern conference was followed by a second such conference held two years later in Chapel Hill, North Carolina.[33] This renewed activity was made possible by the considerable funding made available to theoretical physics after World War II, by the ability of scientists to cross international boundaries, and by the physicists themselves, who recognized the potential in establishing a vibrant community interested in the many aspects of Einstein's theory. Among the most pressing questions that could now be pursued were the existence and properties of gravitational waves. The conference in Chapel Hill led to the broad consensus that gravitational waves are real and that they carry energy.

The pioneering attempts to observe the tiny changes in the length of material rods expected to be produced by such waves began around 1960 with the work of Joseph Weber at the University of Maryland. After ten years of conducting experiments with students and assistants, Weber announced that he had collected convincing evidence for the existence of gravitational waves, which provided the incentive for a series of experiments designed to test his results. None of these experiments, however, confirmed Weber's findings, and they led ultimately to a long and heated controversy. Although a consensus eventually emerged that no gravitational waves had been observed in Weber's experiments, his work had a significant impact on the scientific community. New techniques and methodologies were developed that later constituted the basis of larger measuring systems, including LIGO (Laser Interferometer Gravitational-Wave Observatory).[34]

As well as advances in technology, a second prerequisite for the direct detection of gravitational waves by the LIGO collaboration in 2016 was a better understanding of their astronomical sources. Indeed, "indirect" detection of gravitational waves had already been performed in 1974 by closely observing a binary system of two massive stars that circled each other and slowly lost energy by emitting gravitational waves, with the result that the stars slowed down and moved ever closer to each other. The energy loss observed by Joseph Taylor and Russell Hulse corresponded precisely with theoretical calculations, earning Hulse and Taylor the 1993 Nobel Prize in Physics.

The source of the gravitational waves detected by LIGO today is novel, for it involves the collision of two black holes. Although the concept of black holes resulted from Einstein's equations, their actual existence began to be taken seriously only in the late 1960s, when the existence of singularities were accepted as a generic feature of general relativity and when observations made in the newly established field of radio astronomy (a result of the radio technology developed in the course of radar research in World War II) indicated that such extreme objects really existed at the center of galaxies, including our own.

While the discovery of gravitational waves is now rightly heralded as the confirmation

of a prediction Einstein made one hundred years ago, the likelihood of their existence began to dawn on the scientific community only in the 1950s and 1960s, when the great technological, conceptual, and organizational advances of the renaissance of general relativity laid the groundwork for this discovery and made large-scale, international projects such as LIGO possible. The foundation for this project was laid forty years before the detection of a signal produced by the collision of two black holes, a signal that confirmed ripples in the spacetime structure of the universe and the existence of black holes, which were first speculated about more than a century ago.[35]

NOTES

1. Albert Einstein, et. al. "Diskussion," *Physikalische Zeitschrift* 14 (1913):1262–1266, in CPAE vol. 4, Doc. 18.
2. CPAE vol. 4, Doc. 18, p. 229.
3. Einstein to Karl Schwarzschild, 19 February 1916, CPAE vol. 8, Doc. 194.
4. Albert Einstein, "Näherungsweise Integration der Feldgleichungen der Gravitation," *Königlich Preußische Akademie der Wissenschaften* (Berlin), *Sitzungsberichte* (1916): 688–696, in CPAE vol. 6, Doc. 32. The history of gravitational wave research has been masterfully described by the historian of science Daniel J. Kennefick in his book *Traveling at the Speed of Thought: Einstein and the Quest for Gravitational Waves* (Princeton, NJ: Princeton University Press, 2007).
5. Einstein to Michele Besso, 26 March 1912, CPAE vol. 5, Doc. 377, p. 278.
6. Max Abraham, "Eine neue Gravitationstheorie," *Archiv der Mathematik und Physik* 20 (1913):193–209.
7. Max Abraham, "A New Theory of Gravitation," in *Genesis of General Relativity*, vol. 3, p. 362.
8. Karl Schwarzschild to Einstein, 22 December 1915, CPAE vol. 8, Doc. 169.
9. Karl Schwarzschild to Einstein, 6 February 1916, CPAE vol. 8, Doc. 188.
10. Einstein to Karl Schwarzschild, 19 February 1916, CPAE vol. 8, Doc. 194.
11. Karl Schwarzschild to Arnold Sommerfeld, 17 February 1916, Deutschen Museum, Munich, Archiv NL 89, 059.
12. Albert Einstein, "On the General Theory of Relativity," *Königlich Preußische Akademie der Wissenschaften* (Berlin), *Sitzungsberichte* (1915): 778–86, in CPAE vol. 6, Doc. 21.
13. Einstein to Karl Schwarzschild, 19 February 1916, CPAE vol. 8, Doc. 194.
14. Einstein, "On the General Theory of Relativity."
15. Einstein to Willem de Sitter, 22 June 1916, CPAE vol. 8, Doc. 227.
16. Albert Einstein, "Approximative Integration of the Field Equations of Gravitation," CPAE vol. 6, Doc. 32.
17. Ibid., p. 201.
18. Ibid., p. 209.
19. Ibid.
20. Gunnar Nordström to Einstein, 22 September 1917, CPAE vol. 8, Doc. 382.
21. Gunnar Nordström to Einstein, 23 October 1917, CPAE vol. 8. Doc. 393, p. 390.
22. Albert Einstein, "On Gravitational Waves," CPAE vol. 7, Doc. 1, p. 9.
23. Ibid.," p. 19.
24. Ibid.," p. 23.
25. For the full story, see Daniel J. Kennefick, "Einstein, Gravitational Waves, and the Theoretician's Regress," in *The Cambridge Companion to Einstein*, pp. 270–280; and *Traveling at the Speed of Thought*.
26. Arthur Stanley Eddington, "The Propagation of Gravitational Waves" *Proceedings of the Royal Society of London A* 102 (1922): 268–282, p. 268.
27. Weyl, *Space, Time, Matter*.
28. Arthur Stanley Eddington, *The Mathematical Theory of Relativity* (Cambridge: University Press, 1923), p. 269. *Traveling at the Speed of Thought* is the title of an authoritative book on the history of gravitational waves by Daniel Kennefick.
29. Arthur Stanley Eddington, *Mathematical Theory of Relativity*; and "The Spontaneous Loss of Energy of a Spinning Rod According to Relativity Theory," *Philosophical Magazine* 46 (1923): 1112–1117.

30. Eddington, *Mathematical Theory of Relativity*, p. 95.
31. See the remarks at the end of chapter 5.
32. Blum et al., "Renaissance of General Relativity."
33. Cécile M. DeWitt and Dean Rickles, eds., *The Role of Gravitation in Physics: Report from the 1957 Chapel Hill Conference* (Berlin: Edition Open Access, 2011), http://www.edition-open-access.de/sources/index.html; A. Lichnerowicz and M. A. Tonnelat, eds., *Les Théories Relativistes de la Gravitation*, Colloques Internationaux du Centre National de la Recherche Scientifique, vol. 91 (Paris: CNRS, 1962), p. 475.
34. On the complicated history of Joseph Weber's claims for detection of gravitational waves and the early history of LIGO, see Harry Collins, *Gravity's Shadow: The Search for Gravitational Waves* (Chicago: University of Chicago Press, 2004).
35. Section F is an adaptation of a feature story on the homepage of the Max Planck Institute for the History of Science, "One Hundred Years of Gravitational Waves," by Alexander Blum, Roberto Lalli, and Jürgen Renn, http://www.mpiwg-berlin.mpg.de/en/content/one-hundred-years-gravitational-waves.

7
PHILOSOPHICAL DEBATES ON GENERAL RELATIVITY

Since his student days, Einstein had developed a deep concern for issues of the philosophy and history of science. He read Mach's philosophical-historical critique of classical mechanics as well as his book on the principles of thermodynamics. He was also familiar with Eugen Dühring's history of mechanics and with Ferdinand Rosenberger's book on Newton. He was fascinated by Arthur Schopenhauer and, as a student at the Polytechnic School in Zurich, had attended a course on Kant. Throughout his life, Kant remained a figure with whom Einstein took issue, as the historians Thomas Ryckmann and Michael Friedman have shown.[1] Later, together with his friends from the Akademie Olympia in Bern, he read not only David Hume's *Treatise on Human Nature* and Henri Poincaré's *Science and Hypothesis* but also many other works on epistemology, psychology, and logic. He later recalled in a letter to Michele Besso:

> Together with C. Habicht and Solovine I had regular philosophical reading and discussion evenings in Bern, where we mainly focused on D. Hume (in a rather good German edition). This reading had a significant influence on my development—next to Poincaré and Mach.[2]

General relativity emerged from his concerns for the foundations of physics. His early readings of philosophers such as Hume or philosopher-scientists such as Mach and Poincaré had made him aware of the delicate relation between the fundamental concepts of space and time and experience. In creating special relativity, Hume's empiricism and Poincaré's conventionalism had been helpful because they encouraged Einstein to ascribe new notions of space and time to coordinates that in Lorentz's electrodynamics had no direct physical meaning. His philosophical awareness, however, was much more than a kind of background knowledge enabling him to address concrete physical problems with greater epistemological sensibility than his contemporaries. Rather, the very development of his theories of relativity made it necessary for Einstein to probe this reflective competence and to engage himself in philosophical debates in order to resolve the foundational ambiguities of the emerging theory of general relativity.

The implications of the general theory of relativity constituted a challenge for contemporary philosophical debates, which Einstein actively participated in and shaped for decades to

come. The theory concerned the fundamental notions of space and time, which had been the focus of philosophical epistemology at least since Descartes' time. Given the apparently scarce empirical evidence on which general relativity was based, it raised profound questions about the relation between theory, mathematics, logic, experiment, and reality; its involvement of subtle notions, such as spacetime coincidences, uniqueness, and causality, offered virtually inexhaustible material for the conceptual analysis of the theory and its interpretation; and, finally, its relation to classical and special relativistic physics raised questions about the nature of scientific discoveries and the meaning of progress.

A. MORITZ SCHLICK AND THE MEANING OF COORDINATES

In struggling with the dilemmas of his emerging theory, Einstein had arrived at general relativity only gradually. The philosophical positions he took during this struggle were shaped as much by the state of his current research as by his hopes for the future. When, in the context of the 1913 *Entwurf* theory, he failed to implement the general principle of relativity, he resorted to a particular interpretation of the physical role of coordinates in his theory, constructing the infamous hole argument,[3] which allegedly supported his claim that generally covariant theories cannot be unique. When he finally managed to write down generally covariant field equations in 1915, he had to reconsider this commitment and free himself once more from what now turned out to be a prejudice. Again, philosophical reflections turned out to be of crucial help, in particular the work of the philosopher Moritz Schlick, who had closely followed Einstein's work, trying to interpret it on its own value rather than simply trying to place it into a preconceived system.

Moritz Schlick, whose work Einstein appreciated tremendously, had studied physics himself, obtaining a PhD under Max Planck in Berlin. Later in Zurich, he also studied psychology and philosophy. In 1915 he published a paper on the philosophical meaning of the relativity principle, which Einstein read and admired. On 14 December, Einstein wrote to Schlick:

> Yesterday I received your paper and have studied it thoroughly already. It is among the best that has been written on relativity to date. From the philosophical perspective, nothing nearly as clear seems to have been written on the topic. At the same time, you have complete command of the subject material. I have nothing to criticize about your representations.[4]

Ironically, however, Schlick had started his interpretation of the new relativity theory even before Einstein had achieved his final result. In fact, Schlick's analysis still referred to the preliminary *Entwurf* theory, which Einstein had come to doubt in the days when he wrote the letter to Schlick. Through Schlick's analysis, Einstein was now able to see his own problems in a philosophical mirror. And that mirror helped him to consider his physical problems and, in particular, the intricacies of the hole argument in the wider philosophical context of Schlick's work. Historians of general relativity have long speculated about how Einstein hit upon the escape from the hole argument through his point-coincidence argument.[5] The philosopher and Schlick scholar Fynn Ole Engler was the first to reveal Schlick's role.[6] Schlick, also inspired by his studies of physiology and psychology, had in fact identified spacetime coincidences as a criterion of reality, and his emphasis on them now made it

possible for Einstein to strip the last physical meaning from coordinates and recognize the fault in his hole argument. As mentioned, he had formulated the generally covariant field equations of general relativity before he was able to overcome the hole argument and spot his mistake. But now, after an exchange with Schlick, Einstein recognized that only spacetime coincidences, and not coordinates, matter. This realization helped him to grasp more deeply the physical meaning of what he had achieved.[7] Shortly afterward, on 26 December 1915, Einstein wrote to his friend Paul Ehrenfest:

> Whatever is physically real in events in the universe (as opposed to that which is dependent on the choice of a reference system) consist in *spatio-temporal coincidences* and in nothing else! For ex., the intersection points of two world lines are real, or the statement that they do *not* intersect each other.[8]

In 1917, Schlick published his book *Space and Time in Contemporary Physics: An Introduction to the Theory of Relativity and Gravitation*.[9] Einstein again reacted enthusiastically to Schlick's work when he wrote him on 21 May 1917:

> Time and again I look over your little book and delight in the splendidly clear discourse. I also find the last section, "Links to Philosophy," superb.[10]

In this book, Schlick once again returned to the issue of point coincidences, stressing the origin of this concept in his own psychological and epistemological investigations:

> Upon more careful reflection, one easily finds that we arrive at the construction of physical space and time exclusively by this method of coincidences and in no other way. The spacetime manifold is precisely nothing other than the totality of objective elements defined by this method. This is the result of the psychological-empirio-critical analysis of the space and time concept, and we see that we encounter precisely the significance of space and time that Einstein recognized as alone essential for physics and there gave proper expression.[11]

ESSAY COMPETITION ON GENERAL RELATIVITY

On 10 July 1920 the journal *Scientific American* launched an essay competition on general relativity.[12] The advertisement was sent to newspapers and journals all over the world. The competition was launched by Eugene Higgins with a $5000 prize for the best essay of a maximal length of 3000 words. The fifteen best essays were collected in a book that appeared in 1921: *Einstein's Theories of Relativity and Gravitation. A Selection of Material from the Essays Submitted in the Competition for the Eugene Higgins Prize of $5000*, edited by J. Malcolm Bird.[13] In the introduction, the editor wrote:

> From the beginning we had in view the present volume, and the severe restriction in length was deliberately imposed for the purpose of forcing every contestant to stick to what he considered the most significant viewpoints, and to give his best skill to displaying the theories of Einstein to the utmost advantage from these viewpoints. (p. 7)

The advertisement in *Scientific American* reads:

> An essay of three thousand words is not long enough to lose a reader more than once; if it does lose him it is a failure, and if it doesn't it is a competitor that will go into the final elimination trials for the prize. If we can present, as a result of the contest, six or dozen essays of this length that will not lose the lay reader at all, we shall have produced something amply worth the expenditure of Mr. Higgins' money and our time. For such a number of essays of such character will of necessity present many different aspects of the Einstein theories, and in many different ways, and in doing so will contribute greatly to the popular enlightenment.[14]

Essays from 275 authors from all over the world were submitted to *Scientific American*, including from Canada, France, Denmark, Chile, Cuba, India, South Africa, and the Fiji Islands. The winner was Lyndon Bolton, a collaborator with the British Patent Office. Among the continental European contributors, de Sitter and Schlick were singled out by the editor for the linguistic qualities of their essays:

> Drs. De Sitter and Schlick . . . both showed the ability to compete on a footing of absolute equality with the best of native product.

Schlick's essay, ranked 24 in the competition, ends with the words:

> Einstein's new mechanics are the most marvelous feat of science ever completed. They connect the laws of motion and of gravitation in an inseparable unity, thus furnishing an unexpected wonderful solution of the great problem of gravitation, which has been puzzling science for centuries.

B. PHILOSOPHICAL CHALLENGES OF THE NEW THEORY

Given the impact of philosophical thought on the genesis and formative years of general relativity, it is little wonder that Einstein felt the need to reconsider the various philosophical positions that had marked his pathway. This was one of the reasons for the intensity with which he interacted with philosophers. Another was his hope to learn from what he had already achieved and to extract heuristic clues for his own future pathway. A third motive was the attempt to defend the theory and to resolve some of the puzzles it had created. Fortunately, some of Einstein's philosophical interlocutors, such Moritz Schlick, were themselves deeply familiar with the conceptual and even technical intricacies of the new theory. This was obviously not the case for all of them. When Einstein's close friend Heinrich Zangger sent him some quotations by Henri Bergson, who had published on the philosophy of time, Einstein quipped:

> The philosophical quotes you related to me do not raise an appetite for more. These fellows haven't the least feel for the narrow constraints of the terms they are fumbling around with. When I come and visit you sometime, let's take a closer look at one of them and tear it apart; it would be very interesting for me to experience such a thing with you, but not flaccid Bergson.[15]

The heuristic principles that helped Einstein find his way to the field equations of general relativity were also rooted in philosophical concerns, but, as it turned out, they did not necessarily offer the best means to interpret the final result. This is most evident with the influence of Mach's criticism of Newton's concept of absolute space, which had shaped Einstein's search for a generalization of the relativity principle. While constructing the new theory, Einstein had pursued the aim to conceive all motion as relative to other masses and all inertial properties as being caused by such masses. As it became gradually clear, however, it was hard to reconcile the final version of general relativity with this philosophical demand. As we discussed in chapter 2, his struggle with the implications of Mach's views extended throughout the formative years.

In addition, further interpretational challenges arose that were closely related to the exploration of the newly formulated theory: Does the theory imply the existence of gravitational waves? Does it give rise to a convincing picture of the large-scale structure of the universe? What is the meaning of singularities appearing in its solutions? And how should the energy and momentum of the gravitational field be understood? The discussion of these challenging issues and the gradual realization of a mismatch between the heuristic roots and the conceptual implications of the theory further motivated Einstein's attempts to clarify the fundamental principles of the theory during the formative years.

Clarification became all the more urgent because the empirical evidence supporting the new theory was rather slim, particularly before the British solar eclipse expedition of 1919 confirmed the gravitational light bending and thus supplied a second classical test in addition to the correct prediction of Mercury's perihelion motion. Einstein's engagement with methodological and broader philosophical issues was also a consequence of this precarious situation whereby his theories were often deemed speculative and not sufficiently founded in experience.

C. EINSTEIN'S POPULAR ESSAYS AND THEIR PHILOSOPHICAL CONTEXTS

Between 1916 and November 1918 Einstein wrote several substantial papers on the challenging problems of general relativity, consolidating the theory by exploring its consequences. After the spectacular success of 1919, his engagement with the scientific elaboration of the theory declined. It increasingly gave way to his sustained efforts at explaining the theory both to a wider audience and to the scientific community, efforts that culminated in the Princeton lectures and also several contributions on philosophical issues.

On the occasion of Max Planck's sixtieth birthday in April 1918, Einstein wrote an essay in which he commented on an earlier debate between Planck and Mach about the value of a positivistic philosophy of science, praising the role of intuition in finding the most general laws of physics.[16] He confronted his audience with a puzzle: although he admitted that many equally legitimate theoretical constructions are possible in principle, as claimed by conventionalist philosophy, historical experience has shown that in practice a successful theoretical system is not only unique but also reducible to a few basic principles, which become more and more elementary as science progresses. Describing this circumstance in terms of Leibniz's prestabilized harmony, a conception that presupposes that things have been well, in fact optimally, arranged by the Creator from the beginning, he made no attempt at this point to offer a deeper explanation. Certainly Einstein is also referring here to his own experience with the creation of

general relativity. His reflection on the remarkable way in which science progresses was to have a profound impact on his own further intellectual development: it motivated his quest for an epistemology that could account for what he described as an astonishing circumstance neglected by Mach's philosophy, and it encouraged his attempts to search for a unified field theory.

As the historian and philosopher of science Don Howard has emphasized, the uniqueness of scientific theories was a key subject of contemporary philosophical discussions, in particular in connection with the conventionalist point of view, which held that there are always different ways of describing the empirical facts with a scientific theory.[17] For instance, in his 1915 paper, Schlick wrote:

> It is therefore no contradiction, but lies, rather, in the nature of the matter, that under certain circumstances, several theories may be tried at the same time, in that they achieve indeed a different, but each for itself completely univocal designation of the facts.[18]

Another much discussed philosophical problem of the period was the challenge that both theories of relativity, with their new concepts of space and time, posed to Kantian philosophy, for which space and time were fundamental forms of intuition, given once and for all "a priori" and constitutive of any experience. This position was a crucial part of Kant's answer to the problem of the stability of scientific knowledge in the face of ever-changing sense experiences. The cognitive process itself provides essential building blocks for the architecture of knowledge. But according to general relativity, the concepts of space and time no longer represent a stage, given once and for all, on which the physical events take place but have themselves now become part of the drama. So, was there a role for certain elements of knowledge that are provided not by empirical observation but by the process of cognition? This was a much-debated question during a period in which the new movement later known as logical empiricism struggled with Kantian and neo-Kantian positions, which at that time were dominant in the academic environment.

Next to Schlick, the philosopher Hans Reichenbach became one of Einstein's most important interlocutors. Reichenbach had attended Einstein's course on relativity theory in the summer of 1919 and in his first book undertook a critical revision of Kant's position. In 1920, he wrote to Einstein:

> I must approach you with a big request. In the near future my work on *Relativitätstheorie und Erkenntnis a priori* [Relativity Theory and a priori Knowledge] will be appearing as a booklet (similar to Freundlich's in breadth) with Springer. I would now like to ask for your permission to dedicate the work to you. You know that with this work my intention was to frame the philosophical consequences of your theory and to expose what great discoveries your physical theory have brought to epistemology. By placing your name at the head of the text, I would like to express how greatly philosophy in particular is indebted to you. I know very well that very few among tenured philosophers have the faintest idea that your theory is a philosophical feat and that your physical conceptions contain more philosophy than all the multi-volume works by the epigones of the great Kant. Do, therefore, please

> allow me to express these thanks to you with this attempt to free the profound insights of Kantian philosophy from its contemporary trappings and to combine it with your discoveries within a single system. With this dedication, however, I also want to extend to you my very personal thanks as well, which I owe you. For I was permitted to hear out of your mouth the deepest truths I ever encountered in physics; I shall never forget the great inspiration you gave my intellectual efforts and this text in particular.[19]

Einstein responded:

> I am really very pleased that you want to dedicate your excellent brochure to me, but even more so that you give me such high marks as a lecturer and thinker. The value of the th. of rel. for philosophy seems to me to be that it exposed the dubiousness of certain concepts that even in philosophy were recognized as small change. Concepts are simply empty when they stop being firmly linked to experiences. They resemble upstarts who are ashamed of their origins and want to disown them. Pardon my brevity; my correspondence debts are immense.[20]

As Klaus Hentschel has discussed in detail,[21] in this phase of his thinking Reichenbach still saw his position as a reinterpretation of Kant's belief in the necessity of certain a priori forms of thinking. But in contrast to Kant, he emphasized their historically changing character and did not subscribe to their apodictic nature.

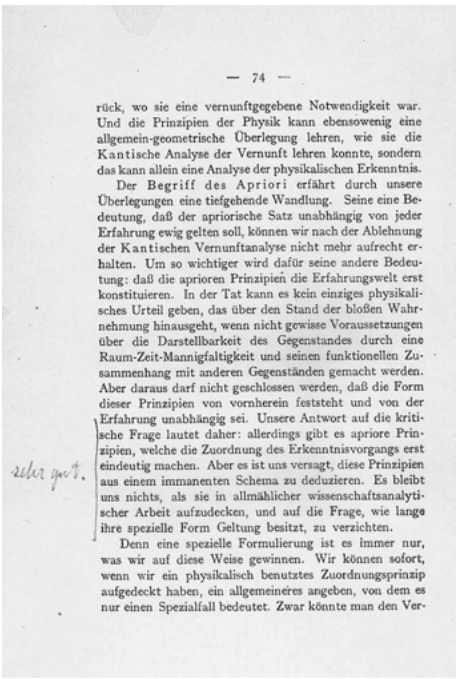
EINSTEIN'S PERSONAL COPY OF REICHENBACH'S FIRST BOOK

In Einstein's personal copy of Reichenbach's first book, he marked a key passage and wrote next to it: "very good!" The passage marked by Einstein ends a longer paragraph in which Reichenbach explains his position with respect to Kant's philosophy. In this passage he distinguished between two meanings of Kant's a priori conditions of knowledge. In their first meaning, these conditions are entirely independent of experience. Clearly, this meaning can no longer be upheld after the new insights of relativity theory. In their second meaning, these conditions nevertheless play a constitutive role for experience. And it is this second meaning to which Reichenbach holds onto when he claimed:

> Indeed, there cannot be a single physical judgment that goes beyond the state of immediate perception unless certain assumptions about the description of the object in terms of a space-time manifold and its functional connection with other objects are made.

Shortly after, Reichenbach wrote the passage that he conceives as his answer to Kant and that Einstein evidently liked so much:

> Our answer to the critical question is, therefore: there are a priori principles that make the coördination of the cognition process unique. But it is impossible to deduce these principles from an immanent scheme. We can detect them only gradually by means of logical analysis of science and must abandon the question of how long their specific form will remain valid.[22]



— 74 —

rück, wo sie eine vernunftgegebene Notwendigkeit war. Und die Prinzipien der Physik kann ebensowenig eine allgemein-geometrische Überlegung lehren, wie sie die Kantische Analyse der Vernunft lehren konnte, sondern das kann allein eine Analyse der physikalischen Erkenntnis.

Der Begriff des Apriori erfährt durch unsere Überlegungen eine tiefgehende Wandlung. Seine eine Bedeutung, daß der aprioriische Satz unabhängig von jeder Erfahrung ewig gelten soll, können wir nach der Ablehnung der Kantischen Vernunftanalyse nicht mehr aufrecht erhalten. Um so wichtiger wird dafür seine andere Bedeutung: daß die aprioren Prinzipien die Erfahrungswelt erst konstituieren. In der Tat kann es kein einziges physikalisches Urteil geben, das über den Stand der bloßen Wahrnehmung hinausgeht, wenn nicht gewisse Voraussetzungen über die Darstellbarkeit des Gegenstandes durch eine Raum-Zeit-Mannigfaltigkeit und seinen funktionellen Zusammenhang mit anderen Gegenständen gemacht werden. Aber daraus darf nicht geschlossen werden, daß die Form dieser Prinzipien von vornherein feststeht und von der Erfahrung unabhängig sei. Unsere Antwort auf die kritische Frage lautet daher: allerdings gibt es apriore Prinzipien, welche die Zuordnung des Erkenntnisvorgangs erst eindeutig machen. Aber es ist uns versagt, diese Prinzipien aus einem immanenten Schema zu deduzieren. Es bleibt uns nichts, als sie in allmählicher wissenschaftsanalytischer Arbeit aufzudecken, und auf die Frage, wie lange ihre spezielle Form Geltung besitzt, zu verzichten.

sehr gut.

Denn eine spezielle Formulierung ist es immer nur, was wir auf diese Weise gewinnen. Wir können sofort, wenn wir ein physikalisch benutztes Zuordnungsprinzip aufgedeckt haben, ein allgemeineres angeben, von dem es nur einen Spezialfall bedeutet. Zwar könnte man den Ver-

FIGURE 17: Key passage in Einstein's personal copy of Reichenbach's first book marked with "very good." H. Reichenbach. *Relativitätstheorie und Erkenntis Apriori* (Berlin: Springer, 1920).

Later, pressed by Moritz Schlick, Reichenbach realized that his views actually amounted to a radical break with Kantian philosophy. In 1924 Reichenbach published his famous book titled *Axiomatik der relativistischen Raum-Zeit-Lehre*, in which he attempted to single out the elementary facts on which the theory is based.[23] He emphasized the role of certain definitions used in physics to establish a connection between theoretical concepts and pieces of reality. Under Schlick's influence he no longer saw these "coordinating definitions" as a substitute for the Kantian a priori structures of thinking, but instead as merely conventional definitions.[24]

Einstein took a very critical stance not only with regard to Kantian philosophy, but also with respect to any such attempt to distinguish within a theory elements that are closer to reality than others. This becomes clear, for instance, from his reaction to the book on the theory of relativity published in 1920 by the philosopher Ernst Cassirer in which the latter attempted to interpret the theory from a neo-Kantian perspective:[25]

> I can understand your idealistic way of thinking about space and time and also believe that one can thereby arrive at a consistent point of view. Not being a philosopher, the philosophical antitheses seem to me more conflicts of emphasis than fundamental contradictions. What Mach calls *connections* [*Verknüpfung*] are for you the ideal names that make experience possible in the first place. You, however, emphasize this aspect of knowledge, whereas Mach wants to have it appear as insignificant as possible. I acknowledge that one must approach experiences with some sort of conceptual tool in order for science to be possible; but I do not think

> that our choice of these tools is constrained *by virtue of the nature of our intellect*. Conceptual systems seem empty to me, if the way in which they are to be related to experience is not established. This seems highly essential to me, even though we often find advantage, in theoretically isolating purely conceptual relations, in order to have the *logically* secure interdependencies come more cleanly to the fore.[26]

As Don Howard has argued, Einstein's own position somewhat resembled that of Pierre Duhem, who emphasized that meaning can only be ascribed to a physical theory as a whole.[27] From the above quotation, it is also apparent, however, that Einstein took the liberty of not subscribing to any specific philosophical position but instead made use of all available arguments if they appeared reasonable to him. Furthermore, in contradistinction to Duhem, Einstein was well aware that the architecture of knowledge was certainly richer than what may be captured by one single theory, as becomes clear from his emphasis of the role of practical geometry, which was discussed in chapter 1. And, as we shall discuss below, he saw more clearly than most of his contemporaries that this architecture was shaped not by the structure of our intellect, but rather by our historically accumulated experience.

In contrast to most contemporary philosophers, Einstein was also more interested in what Reichenbach later called the "context of discovery" rather than in the context of justification of a physical theory.[28] When, at the end of 1919, Einstein wrote for the *Times* in London the article "Space, Time, and Gravitation," he distinguished between two types of physical theories, which he designated as theories of principle and constructive theories.[29] He characterized them in terms of the traditional distinction between the synthetic method, which constructs complex phenomena starting from a simple formalism, and the analytic method, which proceeds from the general properties of natural processes to mathematically formulated criteria that have to be satisfied by the individual physical processes. He argued that the theory of relativity is a theory of principle, like classical thermodynamics, which is based on the principle that there is no perpetuum mobile. In a contribution to the *Berliner Tageblatt* from about the same time, "Induction and Deduction in Physics,"[30] he emphasized that the major advances of science, such as Newton's theory of gravitation or thermodynamics, are due to deduction rather than to induction. He sketched a hypothetico-deductive method according to which theories can never be established as being definitively true but only falsified by empirical findings, a position later developed by Karl Popper into a philosophical system.[31]

With these contributions, which were written for a lay audience, Einstein managed to contextualize his own achievements within the broader historical development of physical theories: Relativity theory, in particular general relativity, may seem to be a theory that is rather different from contemporary attempts, say, to model the motions of electrons in an atom, but if seen in this broader context, it turns out not to be so unusual after all and may in fact be compared to other theories of principle, such as thermodynamics. Also, as in other major advances in physics, it has also been created by deduction rather than by induction, and, most important, general relativity, too, is just a provisional theory that may turn out in the end to be false. What remains is the quest for knowledge, as Einstein explains in the final remark of his article for the *Berliner Tageblatt*:

> I offer the reader in these hectic times a small, objective, passionless reflection because I believe that quiet devotion to the eternal goals that are shared by all civilized men can today serve political reconvalescence better than political meditations and credos.[32]

D. EINSTEIN'S REFLECTIONS ON CONCEPTUAL DEVELOPMENT IN PHYSICS

In papers that are directed at more professional audiences, Einstein offers further reflections on the origins and meaning of relativity theory. He stresses less the methodological and more the epistemological and historical aspects and often also the psychological roots. These papers are therefore more difficult to place within the traditional checkerboard of philosophical positions, from Kantianism, via conventionalism, to positivism, as discussed extensively by Klaus Hentschel.[33] They instead grapple with an understanding of the evolution of physical concepts on the background of historical and psychological reflections, moving closer to what is sometimes discussed today under the label *historical epistemology*.

The first relevant document in this series is the essay "Fundamental Ideas and Methods of the Theory of Relativity, Presented in Their Development," which Einstein completed in early 1920 for publication in the journal *Nature*,[34] responding to the request by Robert Lawson, who later translated Einstein's booklet on the special and the general theory of relativity into English.[35] Lawson asked Einstein to write a short article (3000 words) that would be understandable to nonmathematical readers.[36] Einstein wrote the article but later decided that what he had written was not appropriate for *Nature*. The essay, however, served Einstein as a blueprint for many of his subsequent statements on relativity theory, including the Princeton lectures.

The next publication in which Einstein developed some of the ideas of his unpublished draft was his famous Leiden inaugural lecture, "Ether and the Theory of Relativity," also from 1920. The lecture begins with a psychological account of the origins of concepts, such as that of matter, ether, and action at a distance. The contrast of the latter concept with the intuitive notion that all interactions require contact was, in Einstein's view, an important motivation for the further development of physics because, as he notes: "It is only with reluctance that man's desire for knowledge endures a dualism of this kind." [37]

Einstein proceeds with a discussion of the role of the ether, modeled after a mechanical medium, in explaining an ever-wider set of different physical phenomena, from gravitation to optics to electromagnetism. In the course of this extended use, the ether model was refined, diversified, and not only became even more complex, but was also threatened by contradictions. It was like "Figaro here, Figaro there" for the ether,[38] until it became eventually clear that it could not be coherently construed as a medium with mechanical properties.

Toward the end of his essay, Einstein turns to the role of the ether for general relativity. He argues that Newton's absolute space could also be characterized as an ether since it ascribes physical effects to space. Mach's criticism of this concept could be understood as requiring that this ether not only conditions the behavior of masses, but also that its state is affected by it. Einstein claims: "Mach's idea finds its full development in the ether of the general theory of relativity."[39] This statement translates Mach's views into the language of ether and field theory and thus actually undermines the primacy that Mach had attributed to matter over space, a position to which Einstein had firmly clung but that he now began to rethink.

His essay concludes with the hope that the dualism between matter and ether that had plagued physics for so long might eventually be overcome by a unified field theory of gravity and electromagnetism:

> The contrast between ether and matter would fade away, and, through the general theory of relativity, the whole of physics would become a complete system of thought, like geometry, kinematics, and the theory of gravitation theory.[40]

In other words, Einstein's quest for unification was strongly motivated by his reflections on a conceptual history of physics, reflections that are often neglected when it comes to dealing with Einstein's philosophical position.

About a year later, at the beginning of 1921, Einstein completed a second major essay of the greatest philosophical significance, his famous piece "Geometry and Experience," which is discussed in chapter 1.[41] The reflections he presents in this essay on the meaning of general relativity again include a historical perspective, not only on the development of geometry but also on scientific theories in general. Here we would like to stress that Einstein did not just follow the beaten philosophical paths but boldly strove for his own position, now distancing himself from conventionalism as he had previously distanced himself from Mach's empiricism. A conventionalist viewpoint amounted to the conclusion that one could even adhere to Euclidean geometry forever, unless, as Reichenbach had done, one introduced coordinative definitions ascribing a specific physical significance to certain geometrical notions, thus discarding a holistic interpretation of physical theories.

As noted in chapter 1, Einstein proceeds differently. He distinguishes axiomatic geometry from what he calls practical geometry, which he considers a natural science in its own right, dealing as it does with the placement of rigid bodies. According to Einstein, the conventionalist viewpoint neglects the fact that the concepts of this practical geometry have their origin in an experience that, for the time being (and here Einstein takes a historical perspective), cannot be totally absorbed into an encompassing theoretical construction but remains a layer of knowledge in its own right. As we have also noted in chapter 1, he strongly emphasizes the role that this conception has played in his own creative process: "I attach special importance to the view of geometry which I have just set forth, because without it I should have been unable to formulate the theory of relativity."[42]

This rather sophisticated epistemological position did not square easily with existing philosophical conceptions and was accordingly frequently misunderstood until today. Hermann Weyl, for instance, believed that measurement rods and clocks should be defined within the theory rather than being introduced from the outside. Einstein did not disagree in principle but warned that this was a possibility that may lie far in the future. His position on epistemology therefore had a historical or genetic dimension that was far from the concerns of contemporary philosophy of science.

E. DISCUSSIONS WITH THE PHILOSOPHERS

Einstein not only struggled with the philosophers, but the philosophers also struggled with him. Indeed, as has been argued by historians such as Klaus Hentschel and Don Howard, the development of contemporary philosophy was also shaped by attempts to comprehend, incorporate, or reject Einstein's profoundly new theories, which had shaken up not only classical physics but also challenged the philosophical frameworks that had grown with it.[43] These debates did not just involve Einstein and a few contemporary philosophers, such as Moritz Schlick or Hans Reichenbach, who happened to be interested in and intimately familiar with his work.[44] They also included other contemporary scientists, such

as Hermann Weyl, Erwin Schrödinger, and Arthur Eddington, who were ambitious to continue on Einstein's trail and to achieve further breakthroughs, in particular a unified theory of gravitation and electromagnetism, or a deeper understanding of the riddles of the new quantum theory. Remarkably, practically all early textbooks on general relativity emphasize its significance for philosophy, and even in those that did not, such as the books by von Laue and Thirring, their authors indicate, almost apologetically, that they are not going to engage with philosophical issues, as if doing so was expected of them. We will come back to the early monographs on relativity in chapter 9. And finally, the foundational debates on relativity in the 1920s involved a broad spectrum of physicists, engineers, philosophers, public observers, and amateurs to whom the new theories meant a provocation of common sense, or a welcome occasion to express their hostility to what they considered an unhealthy modernization, or simply another target for their anti-Semitic prejudices.[45]

Nevertheless, the foundational debates on general relativity during these years were among the most fruitful encounters between physics and philosophy in the history of modern science: not only did they have a deep impact on both disciplines, but they even encouraged the hope that philosophy could extract from its close encounter with science a model of rationality that might be generalized beyond it and illuminate society at large.[46] The days of this mutually beneficial exchange and of such high hopes were, however, numbered. The mounting resentment against the liberal, open-minded Weimar culture eventually culminated in the rise to power of the Nazi party.

But even before politics, and eventually the war, ultimately destroyed the conditions for such a courageous intellectual enterprise and dispersed its protagonists, the connections between physics and philosophy had become frail. They were, in fact, hardly supported by any institutional structures and depended primarily on personal contacts and exchanges. These links were thus very much at the mercy of the internal dynamics of different disciplines and schools and also of the vicissitudes of the new foundational theories themselves, first general relativity and then quantum mechanics. By the early 1930s, general relativity had become an increasingly marginal concern of the international physics community, and quantum mechanics, while taking center stage, had become dominated by the interest in its applications and extensions to new domains, such as nuclear physics or quantum field theory, so that philosophical and foundational questions had receded to second-rank priorities.

On the philosophical side, other shifts had meanwhile taken place, in particular the so-called linguistic turn, which had refocused the interest of philosophers such as Moritz Schlick—who had been influenced by the works of Bertrand Russell and Ludwig Wittgenstein—from concerns with logic and science to the fundamental shaping of thought by language. While this turn has had revolutionary consequences for twentieth-century philosophy, it did divert attention from the particulars of science that had been so central to the philosophical investigations of the younger Schlick, perhaps Einstein's most important philosophical interlocutor in the formative years of general relativity. Toward the end of these years, in June 1927, Schlick wrote to Einstein:

> Apart from ethical trains of thought, I am, since a long time ensnared in the new logic (Frege, Russell, Wittgenstein) and I am standing in awe in front of the great labor of thought which has been accomplished therein, and from which I hope nothing less than an entire reform—namely a complete overcoming, making superfluous—of philosophy.[47]

Their ways eventually diverged. In 1930, Einstein informed Schlick:

> Generally speaking, your presentation does not correspond to my way of viewing things, inasmuch as I found your conception, so to speak, too positivistic. Indeed, physics *supplies* relations between sense experiences, but only indirectly. For me *its essence* is by no means exhaustively characterized by this assertion. I put it to you bluntly: Physics is an attempt to construct conceptually a model of the *real world* as well as of its law-governed structure. To be sure, it must represent exactly the empirical relations between those sense experiences accessible to us; but *only* thus is it chained to the latter. . . . You will be surprised by Einstein the "metaphysician." But in this sense every four- and two legged animal is, de facto, a metaphysician."[48]

F. NEGLECTED DIMENSIONS

A disciplinary dimension that is often neglected when recounting the remarkable story of the interaction between physics and reflections on physics in the early twentieth century is that of psychological reflections on the cognitive and creative processes in science. Such reflections were central to the work of Max Wertheimer, the founder of Gestalt psychology. Wertheimer had been in close contact with Einstein since at least 1916 and even had interviewed him to find out more about the creative processes leading to the establishment of special relativity.[49] The founder of genetic epistemology, the Swiss psychologist Jean Piaget, claimed that during a conference in 1928 it was Einstein who inspired him to experimentally investigate the ontogenetic development of the concept of time in children.[50]

Psychological considerations had also been key to the philosophical enterprises of Ernst Mach and Moritz Schlick, in particular their epistemological views, in the same way as they were to Einstein's own reflections on his cognitive processes. But psychology was becoming increasingly established as an independent discipline, "emancipating" itself from philosophy, just as philosophers began to use the term "psychologism" as an accusation against conceiving of logic and the basis of scientific thinking as part of a general investigation of the thinking processes that fell into the domain of psychology.

Finally, and perhaps least surprisingly, witnessing a new science in the making during its formative years did not encourage any deeper historical studies of this process. Not only was the subject of history of science too new, it hardly existed as a field in its own right, so that only recollections on celebratory occasions, some interviews, and the first biographies of Einstein can offer supplementary evidence to the historical sources themselves. Most of the dominant philosophical approaches that were concerned with science were in fact too fundamentally ahistorical in their understanding of science and aloof from its concrete practice, as was observed during the formative years of general relativity. A certain exception was the work of the philosopher Émile Meyerson, who, in his 1925 *The Relativistic Deduction,*[51] placed relativity theory into a broader historical context, even going as far as to compare relativity theory with the systems of Descartes and Hegel, as Einstein emphasized in the draft of a review of Meyerson's work.[52]

Approaches that did pay attention to this practice rather than just to cleaned-up, textbook versions of science remained at the margins of contemporary discussions. The reference is, of course, to the pioneering studies of Ludwik Fleck on thought collectives

and the emergence of scientific facts that deal with biological and chemical matters rather than with physics.[53] Nevertheless, Fleck was aware of the epistemological provocation his studies entailed and vainly attempted to publish them in the prestigious series of philosophical monographs edited by Moritz Schlick.[54]

These are just some glimpses at the rich yet fragmented landscape of what one might describe as the epistemological resources available to Einstein and his contemporaries during the formative years of general relativity. Einstein himself was so involved with many of the contemporary philosophical discussions and, in writings and correspondence, was so outspoken about philosophical positions and their relation to his science that he himself is often described as an active philosopher of science. Yet it has proven notoriously difficult to pinpoint his exact position within the stream of an ongoing debate in which he often seemed to vacillate, changing or adapting his viewpoint, not of course to accommodate the mainstream or current trends, but rather to motivate, to articulate, or even to comprehend his own scientific undertakings. This difficulty in identifying Einstein with one of the main philosophical schools, in finding the right drawer for his position, has led to his characterization as an "opportunist," someone who eclectically picks and chooses among the available resources to defend his own point of view according to the needs of his science. But if picking and choosing among the treasures of a tradition really is an opportunistic characteristic, then all great thinkers of the past have been and had to be opportunists. The question is less about what they picked and more about what they actually did with their choices.

What Einstein made of the philosophical debates in which he participated was not primarily philosophy, but it was not just science either. This broader reflection was all the more needed because in these formative years the empirical evidence in the sense of novel effects that were predicted by the theory and could be observationally confirmed was still rather weak. And since these were the formative years and the theory itself was still in flux, so naturally was Einstein's thinking about his science. It was Moritz Schlick who perhaps best captured the inseparability of the physical and philosophical aspects of general relativity:

> The doctrine of relativity is primarily a physical theory. But whoever, for this reason, denies its philosophical character and the philosophical impact of the theory (as one has occasionally done) misconceives that the physical and philosophical perspectives can in no way always be rigorously separated from one another, that both merge into each other once they turn to the treatment of the highest, most general foundational concepts of physics. It is due to its philosophical content that relativity has ignited such an overly great interest among the intellectual community: as a matter of fact one clearly feels how deeply one intervenes in the human habits of thinking; and humans like to be shocked.[55]

NOTES

1. Thomas Ryckman, "'A Believing Rationalist': Einstein and 'the Truly Valuable' in Kant," in *The Cambridge Companion to Einstein*, pp. 377–397; Friedman, "Space, Time, and Geometry."
2. Einstein to Michele Besso, 6 March 1952, AEA 7-402.
3. See, for example, Gutfreund and Renn, *The Road to Relativity*, p. 25.
4. Einstein to Moritz Schlick, 14 December 1915, CPAE vol. 8, Doc. 165.

5. On the role of correspondence with Paul Hertz, see Don Howard and John D. Norton, "Out of the Labyrinth? Einstein, Hertz, and the Göttingen Answer to the Hole Argument," in *The Attraction of Gravitation*, pp. 30–62; and Erich Kretschmann, "Über die prinzipielle Bestimmbarkeit der berechtigten Bezugssysteme beliebiger Relativitätstheorien (I)," *Annalen der Physik* 48 (1915): 907–942.
6. Fynn Ole Engler and Jürgen Renn, "Hume, Einstein und Schlick über die Objektivität der Wissenschaft," in *Moritz Schlick—Die Rostocker Jahre und ihr Einfluss auf die Wiener Zeit*, ed. F. O. Engler and M. Iven, Schlickiana 6 (Leipzig: Universitätsverlag, 2013), pp. 123–156.
7. Engler and Renn, "Hume, Einstein und Schlick."
8. Einstein to Paul Ehrenfest, 26 December 1915, CPAE vol. 8, Doc. 173.
9. Moritz Schlick, *Raum und Zeit in der gegenwärtigen Physik: zur Einführung in das Verständnis der allgemeinen Relativitätstheorie* (Berlin: Springer, 1917). Translated into English in 1920 by Oxford University Press.
10. Einstein to Moritz Schlick, 21 May 1917, CPAE vol. 8, Doc. 343.
11. Moritz Schlick, *Space and Time in Contemporary Physics: An Introduction to Theory of Relativity and Gravitation* (New York: Oxford University Press, 1920), p. 82.
12. See Fynn Ole Engler, "Moritz Schlicks Beitrag zum Einstein-Wettbewerb des Scientific American," in *Moritz Schlick—Stationen. Dem Philosophen und Physiker Moritz Schlick zum 125. Geburtstag*, ed. F. Stadler und H. J. Wendel (Wien: Springer, 2009), pp. 281–291.
13. New York: Scientific American Publishing and Munn & Co., 1921.
14. *Scientific American. The Weekly Journal of Practical Information*, 10 July 1920, vol. 123, no. 28.
15. Einstein to Zangger, after 27 December 1914, CPAE vol. 10, Doc. 41a, pp. 13–14.
16. Albert Einstein, "Motives for Research," CPAE vol. 7, Doc. 7.
17. Don Howard, "Einstein and the Development of Twentieth-Century Philosophy of Science," in *The Cambridge Companion to Einstein*, pp. 354–376.
18. Moritz Schlick, "Die philosophische Bedeutung des Relativitätsprinzips," *Zeitschrift für Philosophie und philosophische Kritik* 159 (1915): 129–175, p. 149; translation from Don Howard, "Relativity, *Eindeutigkeit*, and Monomorphism," in *Origins of Logical Empiricism*, ed. R. N. Giere and A. W. Richardson (Minneapolis: University of Minnesota Press, 1996), pp. 115–164.
19. Hans Reichenbach to Einstein, 15 June 1920, CPAE, vol. 10, Doc 57.
20. Einstein to Hans Reichenbach, 30 June 1920, CPAE, vol. 10, Doc. 66.
21. Klaus Hentschel, *Interpretationen und Fehlinterpretationen der speziellen und der allgemeinen Relativitätstheorie durch Zeitgenossen Albert Einsteins* (Basel: Birkhäuser, 1990), section 4.1.
22. Hans Reichenbach, *Theory of Relativity and A Priori Knowledge* (Berkeley: University of California Press, 1965 [1920]), pp. 77–78.
23. Hans Reichenbach, *Axiomatization of the Theory of Relativity* (Berkeley: University of California Press, 1969. Originally published as *Axiomatik der relativistischen Raum-Zeit-Lehre* in 1924.
24. Hentschel, *Interpretationen und Fehlinterpretationen*, section 5.1.
25. Ernst Cassirer, *Zur Einstein'schen Relativitätstheorie: erkenntnistheoretische Betrachtungen* (Berlin : Cassirer Verlag, 1920).
26. Einstein to Ernst Cassirer, 5 June 1920, CPAE vol. 10, Doc. 44.
27. Howard, "Einstein and the Development of Twentieth-Century Philosophy of Science."
28. Hans Reichenbach, *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge* (Chicago: University of Chicago Press, 1938).
29. Albert Einstein, "Time, Space; and Gravitation," *Times* (London), 28 November 1919, pp. 13–14; CPAE vol. 7, Doc. 26.
30. Albert Einstein, "Induction und Deduction in der Physik," *Berliner Tageblatt*, 25 December 1919, morning edition, supplement 4, p. 1; "Induction and Deduction in Physics," CPAE vol. 7, Doc. 28.
31. Karl Popper, *The Logic of Scientific Discovery* (London: Hutchinson, 1959 [1934]), chap. 1.
32. Albert Einstein, "Induction and Deduction in Physics," CPAE vol. 7, Doc. 28, p. 109.
33. Hentschel, *Interpretationen und Fehlinterpretationen.*
34. CPAE vol. 7, Doc. 31.
35. Albert Einstein, *Relativity: The Special and the General Theory: A Popular Exposition by Albert Einstein*, trans. R. W. Lawson (London: Methuen, 1920).
36. Robert Lawson to Einstein, 26 November 1919, CPAE vol. 9, Doc. 177.
37. CPAE vol. 7, Doc. 38, p. 163.

38. From the aria "Largo al Factotum," *The Barber of Seville*, by Gioachino Rossini.
39. Ibid., p. 176.
40. Ibid., p. 181.
41. Albert Einstein. "Geometry and Experience," in *Sidelights of Relativity*, ed. A. Einstein (London: Methuen, 1922), pp. 17–18.
42. Ibid., p. 18.
43. Howard, "Einstein and the Development of Twentieth-Century Philosophy of Science"; Hentschel, *Interpretationen und Fehlinterpretationen,* chap. 4.
44. Fynn Ole Engler, "Moritz Schlick und Hans Reichenbach über die *Eindeutigkeit* der Zuordnung, die Gründe diese aufzugeben und die heuristische Stärke eines Empirismus mit begriffskonstitutiven Prinzipien," in *Moritz Schlick—Leben, Werk und Wirkung*, ed. F. O. Engler and M. Iven (Berlin: Parerga, 2013), pp. 131–191.
45. Milena Wazeck, *Einstein's Opponents: The Public Controversy about the Theory of Relativity in the 1920s* (Cambridge: Cambridge University Press, 2014). Originally published in German as *Einsteins Gegner: Die öffentliche Kontroverse um die Relativitätstheorie in den 1920er Jahren* (Frankfurt: Campus Verlag, 2009). See also Hubert Goenner, "The Reaction to Relativity Theory I: The Anti-Einstein Campaign in Germany in 1920," *Science in Context* 6:1 (1993): 107–133.
46. Fynn Ole Engler and Jürgen Renn, *Gespaltene Vernunft. Einstein—Fleck—Schlick* (Berlin: Matthes & Seitz, forthcoming).
47. Moritz Schlick to Einstein, 5 June 1927, EA 7-402.
48. Einstein to Moritz Schlick, 28 November 1930, EA 21-603.
49. See Max Wertheimer, *Productive Thinking* (New York: Harper, 1943); Peter Damerow, "Albert Einstein e Max Wertheimer," in *Percorsi della scienza storia testi problemi*, ed. G. Pisent and J. Renn (Padova: Il poligrafo, 1994), pp. 43–60.
50. Jean Piaget, *The Child's Conception of Time* (New York: Basic Books, 1969 [1946]), p. ix.
51. Émile Meyerson, *The Relativistic Deduction: Epistemological Implications of the Theory of Relativity* (Dordrecht: Reidel, 1985). First published as *La déduction relativiste* (Paris; Payot, 1925).
52. Hentschel, *Interpretationen und Fehlinterpretationen*, p. 463.
53. Ludwik Fleck, *Genesis and Development of a Scientific Fact* (Chicago: University of Chicago Press, 1979 [1935]).
54. Engler and Renn, *Gespaltene Vernunft*.
55. Moritz Schlick, "Die Relativitätstheorie in der Philosophie," in *Verhandlungen der Gesellschaft deutscher Naturforscher und Ärzte. 87. Versammlung zu Leipzig, Hundertjahrfeier vom 17. bis 24. September 1922*, ed. A. Wittig (Leipzig: Vogel, 1923), p. 58. Authors' translation.

8
THE QUEST FOR A UNIFIED FIELD THEORY

With the completion of the general theory of relativity there were two distinct dynamical field theories on the physical arena, which Einstein attempted to unify.[1]

In the nineteenth century, James Maxwell had created the first successful field theory. Maxwell's equations tell us how the electric and magnetic components of the electromagnetic field change in space and time, how they are generated by electrical charges and currents, and how electromagnetic waves propagate. The motion of charged particles is determined by the field that they produce. The three components of the electric field vector and the magnetic field vector are derived from the electromagnetic potential, which can be described by a four-vector, the components of which are functions of spacetime coordinates. With the advent of special relativity, Maxwell's equations have been incorporated into the four-dimensional geometrical formulation of spacetime in special relativity.

In the relativistic theory of gravitation, Einstein succeeded in transforming Newton's instantaneous action-at-a-distance gravitational force between two masses into a field theory of gravitation. One difference between these two fields is that electromagnetism is represented by a physical field that has no effect on the geometrical properties of spacetime, whereas the gravitational field is described by the non-Euclidean geometry of spacetime. The basic concept is the line element, which measures the distance (rather the square of the distance) between two adjacent points in spacetime. The line element is a linear combination of products of the coordinate displacements between the two points (eq. 55, p. 224 [64]). The coefficients $g_{\mu\nu}$ are ten independent components of the symmetric metric tensor. They characterize the local geometry of spacetime and, at the same time, serve as components of the gravitational potential. Thus, spacetime is a dynamic entity; its geometry changes with changes in the gravitational field. In Riemannian geometry this line element is an invariant. All observers applying standard measuring rods and clocks, whatever their state of motion and whatever coordinate system they use, will agree on its value.

Although there was no compelling argument that required a geometric representation of the electromagnetic field of the kind of the gravitational field, Einstein's theory of gravitation was perceived by himself, and by others, as partial and incomplete and in need of supplementation by a unified theory, which would include the electromagnetic field on

the same footing. In his Nobel Prize address, which he delivered to the Nordic Assembly of Naturalists at Gothenburg in July 1923 (Einstein was not present at the Nobel Prize award ceremony in 1922), Einstein defined this as a goal to be achieved:

> The mind striving after unification of the theory cannot be satisfied that two fields should exist which, by their nature, are quite independent. A mathematically unified field theory is sought in which the gravitational field and the electromagnetic field are interpreted only as different components or manifestations of the same uniform field, the field equations where possible, no longer consisting of logically mutually independent summands.[2]

A related but independent problem, which in those days attracted the attention of scholars, concerned the coexistence of fields and material particles and addressed the question of the nature and origin of matter. Electromagnetism and gravitation were both dualistic theories in the sense that the sources of the field are charges or masses not accounted for by the theories themselves. In both theories, the motion of these particles and charges was described by equations of motion in addition to the field equations. Thus, the goal of unification posed a double challenge: to represent gravitation and electromagnetism as one geometric field in which matter and its motion would be deduced from the field equations themselves. Such a unification scheme was expected to account for the basic properties of matter, in particular for the nature of the two particles known at that time: the electron and the proton. A theory of this kind would explain the microscopic properties of elementary particles, the constituents of atoms, and at the same time would account for the macroscopic phenomena constituting the fabric of the universe. These were the goals pursued by Einstein and an increasing number of his contemporaries.

A. THE NATURE AND ORIGIN OF MATTER

In 1912, Gustav Mie had already proposed a unified description of charged particles and electromagnetic fields. His idea was to search for nonlinear formulations of Maxwell's equations, which would allow for solutions with very high intensity in a small region around a point in space. With an appropriate equation of motion, such a region in space could be interpreted as a particle. Keeping in mind that the only known particles were the electrically charged electron and proton (the neutron was discovered only in 1932), such a scheme would imply that the nature of material particles is purely electromagnetic.

Mie's idea appealed to physicists and mathematicians, specifically to David Hilbert. Einstein had also adopted it for a short time. When he submitted his theory of general relativity to the Prussian Royal Academy of Sciences on 4 November 1915,[3] it was still not fully covariant: there was a restriction on the admissible coordinate transformations. To satisfy energy-momentum conservation, the determinant of the metric tensor could not be equal to one. In a subsequent paper, submitted a week later on 11 November, Einstein was able to remove this obscure restriction by "introducing an admittedly bold additional hypothesis on the structure of matter."[4] The assumption was that only those fields occurring as sources of gravitation are electromagnetic fields, so that ultimately all matter can be reduced to the latter. Only in the final paper, submitted on 25 November,

did Einstein find a way of removing this assumption on the nature of matter and write down an equation applicable to all kinds of matter.

In his first comprehensive summary of the general theory of relativity, which was published in 1916, Einstein alluded to the challenge of bringing electromagnetism and gravitation into the framework of a single theory. He left open the question whether the combination of gravitation and electromagnetism would lead to a new theory of matter, which Gustav Mie had tried to build on electromagnetism alone:

> In particular it may remain an open question whether the theory of the electromagnetic field in conjunction with that of a gravitational field furnishes a sufficient basis for the theory of matter or not. The general postulate of relativity is unable on principle to tell us anything about this. It must remain to be seen, during the working of the theory, whether electromagnetics and the doctrine of gravitation are able in collaboration to perform what the former by itself is unable to do.[5]

This hope constituted the nucleus of Einstein's lifelong search for a unified field theory. It was rooted not just in a general quest for unification, as is often assumed, but also in the specific intuition, expressed in this quotation, that the "double dualism" of the coexistence of two fundamental fields, electromagnetism and gravity, on the one hand, and the dualism of particles and fields, on the other, may complement each other and give rise to an encompassing, nondualistic field theory. Although Einstein was quite aware from the beginning that he did not have a strong heuristic principle such as the equivalence principle to guide him in this endeavor, the challenge represented by this double dualism was specific enough to be taken on as a mission in life, in particular because it remained disregarded by most other physicists.

In 1919, in the reports of the Prussian Academy, Einstein published an article titled "Do Gravitational Fields Play an Essential Role in the Structure of the Elementary Particles of Matter?"[6] In this article he argues that there are indications for the idea that the electrical elementary building blocks of atoms are held together by gravitational forces. The stability of the electron had been a long-standing challenge for electromagnetic theorists, among others, for Mie. As we have discussed above, Einstein had, in late 1915, also thought about the possibility that gravitational fields play a role in the constitution of matter. Now he went back to consider a modification of the field equation of general relativity, inspired by the search for such a connection between gravity and the constitution of matter. The fact that, in order to find a satisfactory solution to the cosmological problem, he had already modified these field equations by introducing the cosmological constant may have provided a further stimulus for exploring possibilities of such modifications. Einstein had formulated the original field equation with an eye to how energy and momentum conservation works out in special relativity, that is, in the absence of gravitational fields. But if gravitational fields are now assumed to be always present because they are part of the constitution of elementary particles, the constraint that energy conservation should take this particular form in the limit of vanishing gravitational fields could be abandoned, thus opening up a space for the modification of the field equation. He hoped that the freedom achieved would enable a satisfactory modification to be found to allow him to account for

the stability of elementary particles, but he had to conclude that "the problem of the constitution of the elementary quanta cannot yet be solved on the immediate basis of the given field equations."[7]

In *The Meaning of Relativity*, Einstein referred briefly to this issue. He criticized an ad hoc inclusion of electricity in the general theory of relativity and advocates a theory that would truly unify the electromagnetic and gravitational fields:

> This inclusion of electricity in the scheme of the general theory of relativity has been considered arbitrary and unsatisfactory by many theoreticians. Nor can we in this way understand the equilibrium of the electricity, which constitutes the elementary electrically charged particles. A theory in which the gravitational field and the electromagnetic field do not enter as logically distinct structures would be much preferable.[8]

The summary of Einstein's fifth Princeton lecture, published by Professor Adams in the *New York Times* on 14 May 1921, indicates that Einstein discussed in greater detail there than in the published version the electrical structure of matter assumed by certain researchers. Adams summarized the serious difficulty emphasized in this assumption, as pointed out by Einstein:

> For if we regard a portion of electricity which must be supposed to have a finite although it may be a very small volume, the repulsion between elements of that volume requires some unknown force in order that it may hold together. In order to get around this difficulty, Poincaré assumed the existence of a kind of pressure in the universe of sufficient magnitude to balance the electrostatic repulsion between the elements of electricity. Professor Einstein showed that this assumption of pressure throughout the universe is wholly consistent with the general theory of relativity, although it does not follow as a consequence of the theory.[9]

B. FIRST ATTEMPTS TO UNIFY GRAVITATION WITH ELECTROMAGNETISM

The challenge of generalizing Einstein's theory of gravitation generated interest among researchers from the beginning of the formative period of general relativity. The pioneers in this effort were Hermann Weyl, Theodor Kaluza, and Arthur Eddington. Einstein was very active in this endeavor, though in the first years he was involved primarily by reacting to the work of others through correspondence and comments published mainly in the *Proceedings of the Prussian Academy of Sciences*. His own work on this project started in 1925 and continued until the end of his life, when most of the other players had shifted to other topics.

At the time of the Princeton lectures, Einstein was aware of the work of Weyl and Kaluza, but not yet of Eddington's. In *The Meaning of Relativity* he refers to them in one sentence:

> A theory in which the gravitational field and the electromagnetic field do not enter as logically distinct structures would be much preferable. H. Weyl, and recently

> Th. Kaluza, have put forward ingenious ideas along this direction; but concerning them I am convinced that they do not bring us nearer to the true solution of the fundamental problem. I shall not go into this further." (p. 258 [98])

In spite of this remark, at that time and in later years, Einstein expressed great interest in their work, as we show below.

The first attempt to combine gravitation and electromagnetism in a unified field theory was proposed by Hermann Weyl:

> Lately I succeeded, I believe, in deriving electricity and gravitation from a single common source. A completely determinate action principle results which in the electricity-free case leads to your gravitation equations; in the gravitation free one, on the other hand, it yields equations that agree in *first approximation* with Maxwell's.[10]

Weyl hoped that Einstein would submit this paper to the *Proceedings of the Prussian Academy of Sciences*. Only members of the academy could submit their papers or endorse the publication of papers by nonmembers.

Weyl's proposal is based on an extension of Riemannian geometry, which makes use of the affine connection (see p. 20), which had been introduced by Levi-Civita. In Euclidean geometry, one can compare the length and direction of two vectors at different points in space.

In the new geometry, the length of a vector changes as it is parallel-transported along a specific curve, and the amount of change depends on the curve. As a result, the line element between two adjacent points in spacetime is not invariant, as it is in Riemannian geometry. The metric tensor is determined up to a scale factor (a gauge). This scale factor is a linear combination of the coordinate displacements between the two points. The coefficients in this expression are functions of spacetime coordinates and may be identified as the components of the electromagnetic potential. If these functions are equal to zero, there is no electromagnetic field and the previous geometry is valid. In the general case, fourteen coefficients appear—ten of them describing gravitation and four electromagnetism.

When Einstein received Weyl's manuscript, he responded with reserved enthusiasm:

> Your paper has arrived. *It is a first class stroke of a genius*. However, I have not been able to settle my measuring-rod objection yet.[11]

And four days later, urging him to send immediately an abstract of his paper for the Academy proceedings, Einstein wrote:

> Your chain of reasoning is so wonderfully self-contained. . . . Except for the agreement with reality, it is in any case a grand intellectual achievement.[12]

Einstein's criticized Weyl's theory because it implied that the measuring devices—rods and clocks—lose their invariance and depend on the history of their movement in spacetime. Atoms may act as standard clocks and the frequency of the emitted light may be used to measure time intervals. If this frequency were to depend on history, then identical atoms might emit light at slightly different frequencies and we would not observe sharp spectral lines.

After some correspondence discussing this controversial issue, Einstein submitted Weyl's paper to the academy, and it was published with Einstein's critical comment and with Weyl's response. In the paper, Weyl emphasizes the success of his theory to fully geometrize all of physics:

> According to the theory that now takes shape, both classes of phenomena spring from the same source, and in fact *we cannot in general make any arbitrary separation of electricity from gravitation*. In this theory *all physical quantities have a meaning in world geometry*.[13]

Referring to his conclusion that Weyl's theory implies that chemical elements would not emit spectral lines of well-defined frequency, Einstein states that

> because this is not the case, the basic hypothesis of the theory is not acceptable, to my mind, though its depth and audaciousness must fill every reader with admiration.

Responding to Einstein's criticism, Weyl argued that the assumption that line elements can be measured directly with rods and clocks is problematic, and that rods and clocks can be dispensed with as measuring devices. The line element and the components of the metric tensor can be measured with light signals, up to a constant factor reflecting the freedom to choose a global unit length. He pointed out that the line element, to a very good approximation, will almost always be equal to that in general relativity. But even if one assumes that clocks measure the line element along their world line, sharp spectral lines would still exist because "the influence of any stormy prehistory will quickly fade away." Einstein was not convinced. He continued to express his admiration for theoretical mathematical elegance of the theory, rejecting it as a description of physical reality.

Einstein shared his views with colleagues. To Zangger he wrote about Weyl's attempt:

> It is my firm conviction that it is a wrong path; but he believes firmly in it and only God knows the truth.[14]

He expressed the same attitude in a letter to Eddington:

> I do not consider Weyl's theory (metric interpretation of the electromagnetic potential) correct. It seems to me incompatible with the fact that measuring rods and clocks exhibit behavior independent of their prehistories.[15]

When Weyl included his theory as a chapter in the third edition of *Space, Time, Matter*, Einstein wrote to Ehrenfest:

> Weyl has now added his electromagnetic theory to the new edition of his textbook, unfortunately, so his admittedly very ingenious nonsense will make its way into the cerebra. But I console myself that the sieve of time will do its work here as well.[16]

In the letter to Zangger, Einstein used even harsher terms:

> He is a very remarkable mind but a little removed from reality. In the new edition of his book, he made a complete mess of relativity, I think—God forgive him. Perhaps he will eventually realize that, that for all his keen perception, he has shot wide of the mark.[17]

Although Weyl's theory did not achieve its goal, the effort was not wasted and it left a lasting impact. Weyl formulated his theory as a gauge field theory. It was a precursor to the central role of gauge symmetry and gauge field theories in modern physics.

GAUGE FIELD THEORIES

The fundamental interactions in nature are described by field equations, which can be derived from a mathematical object known as the action, which may be invariant under certain transformations at every point in spacetime. A group of such transformations forms the gauge symmetry of the particular field theory. Quantities like the electromagnetic potentials are themselves not physically observable. Different values (as functions of spacetime coordinates) may lead to the same observed quantities. Such physically equivalent field configurations are characterized by a specific choice of gauge. The most fundamental theory of microscopic interactions, the Standard Model of Elementary Particles, is based on such gauge theories, which are a cornerstone of modern physics. The underlying concepts first emerged in the context of classical electromagnetism and Einstein's theory of gravitation. The term "gauge" was introduced to mathematics and physics by Weyl in the context of his attempt to unify electromagnetism with gravitation. About ten years later, in 1929, already in the context of quantum mechanics, Weyl proclaimed the invariance properties of a field theory as a general principle of physics. He called it *Eichinvarianz* in German and "gauge invariance" in English.

Arthur Eddington was another pioneer in the quest for unification. In his *Space, Time and Gravitation*, he articulated the need to meet this challenge:

> Yet we cannot rest satisfied until a deeper unity between the gravitational and electrical properties of the world is apparent. The electron, which seems to be the smallest particle of matter, is a singularity in the gravitational field and also a singularity in the electrical field. How can these two facts be connected? The gravitational field is the expression of some state of the world, which also manifests itself in the natural geometry determined with measuring appliances; the electric field must also express some state of the world, but we have not as yet connected it with natural geometry. May there not still be unnecessary assumptions to be removed, so that a yet more comprehensive geometry can be found in which gravitational and electrical fields both have their place?[18]

Eddington devoted a whole chapter (chapter 11) to an analysis of Weyl's suggestion of a "more comprehensive geometry" that could achieve this goal. A year later, Eddington ventured his own proposal to generalize Riemannian geometry. His goal was broader than the unification of the two fields. Like Einstein and others, Eddington regarded the energy-momentum tensor in the field equations of Einstein's theory of general relativity as provisional. In a truly unified theory, the source term would automatically arise as some aspect of the free-space field equations. He also shared the hope that an improved fundamental theory would explain why the two elementary particles known at that time (proton and electron) have such different masses.

Instead of starting with the line element and the metric tensor g and using them to derive the components of the affine connection Γ, he started from the affine connection as the fundamental structure. Eddington assumed the affine connection to be symmetric in its lower (covariant) indexes, because it seemed plausible that the result of infinitesimally transporting one vector toward another should produce the same result as transporting the second toward the first. From the affine connection, one can generate the Riemann tensor and thereof the Ricci tensor (as in *The Meaning of Relativity*, pp. 235–237 [75–77]. This procedure may lead to a non-symmetric Ricci tensor. Eddington identified the symmetric components with the gravitational field and the asymmetric components with the electromagnetic field.

Einstein welcomed this approach and, as we shall see, he thought that it was more promising than Weyl's theory. The preference of Γ over g was very much in line with his thinking. Years later, toward the end of his life, he stressed the role of Levi-Civita's notion of the affine connection (which had been developed after the completion of general relativity), rather than Riemann's concept of a metric, as the appropriate mathematical representation of the key insight of general relativity, namely, the elimination of rigid space:

> It seems to me that Levi-Civita's most important contribution lies in the following theoretical discovery: the most essential theoretical accomplishment of general relativity, namely the elimination of "rigid" space, i.e. of the inertial system, is only indirectly connected with the introduction of Riemannian metric. The immediately essential conceptual element is the "displacement field" (Γ^{l}_{ik}), which expresses the infinitesimal displacement of vectors.[19]

What remained was to find a physically acceptable field equation for the affine connection Γ. This is where Einstein stepped in. On the steamer *Haruna Maru*, which took him from Japan back to Berlin, Einstein wrote a paper comparing the ideas of Weyl and Eddington and suggesting a way to derive the field equations. It took him four days in February 1923 to write this paper, which was later submitted for publication in the proceedings of the Prussian Academy. Einstein discards Weyl's approach as not being the way to arrive at a physically useful theory. He proposes to obtain the field equations from a variational principle and remarks that it is not clear if this theory sheds any light on the charged particles. In conclusion,

> the above analysis demonstrates that Eddington's general idea, combined with Hamilton's principle, leads to a theory virtually free of arbitrariness that does justice

> to our current knowledge about gravitation and electricity and unifies the two types of field in a truly perfect way.[20]

A short time before he completed this paper he wrote to Niels Bohr:

> I now believe I finally understand the connection between electricity and gravitation. Eddington came closer to the gist of it than Weyl.[21]

About half a year later, Einstein summarized his ideas in a short paper, "The Theory of the Affine Field."[22] There he concluded that the theory "brings no enlightenment on the structure of electrons." By that time he knew that the theory could not account for the electron-proton mass difference and that it did not have singularity-free solutions that could be interpreted as electrons.

Riemannian Geometry in Five Dimensions. The mathematician Theodor Kaluza proposed a different approach. He remained within the realm of Riemannian geometry but added a fourth space dimension, rendering a five-dimensional spacetime. The new dimension was of a unique nature. Moving along that direction, at any point in spacetime, would eventually bring you back to the starting point. Kaluza's model was referred to by Einstein as a "five-dimensional cylindrical world."[23] In general relativity, ten independent functions of coordinates are needed to describe the spacetime geometry at every point. Fifteen such functions are required to describe the local geometry in five dimensions. Kaluza showed that four of the additional functions could represent the electromagnetic vector potential from which the electric and magnetic fields are derived. The meaning of the fifteenth component of the metric g_{55} remained unclear. Kaluza argued that this component is not truly independent; rather, it is indirectly determined through the field equation. In later works, this component was interpreted as a new scalar field that may be given an independent physical meaning. One proposal in this context was due to Einstein himself: it could represent the global pressure, conjectured by Poincaré, which keeps charged particles from disintegrating because of the electrical repulsion between parts of their charge.

At first, Einstein was impressed:

> At first glance your idea has great appeal for me. It seems to me to have decidedly more promise from the physical point of view than the *mathematically* probing exploration by Weyl, because it starts with the electric *field* and not with the physically meaningless—I am convinced—four *potential*.[24]

He indicates in this letter that he would gladly present the paper to the proceedings of the academy "provided that upon reading a more thorough exposition I do not see what I would consider to be some crushing objection that cannot be resolved."

A week later, Einstein suggested to Kaluza to show that the five-dimensional geodesic lines in his theory could be interpreted as the world lines (trajectories in four dimensional spacetime) of electrically charged particles under the simultaneous influence of a gravitational and an electromagnetic field: "*If you could show with the precision established by our*

empirical knowledge that this is true, I would be as good as convinced of the correctness of your theory."[25]

However, on further reflection on Kaluza's manuscript, Einstein found a serious problem when he estimated the influence of the gravitational field on the motion of an electron and discovered that it exceeded by many orders of magnitude any physically reasonable magnitude of this influence.[26] Einstein could not support the publication of Kaluza's article, and their correspondence ceased for about two years, until Einstein wrote to Kaluza that he had had second thoughts about rejecting his paper two years ago and indicated that he would be willing to present it to the academy after all. He argued: "Your route seems to me, in any event, to have more in its favor than the one H. Weyl has taken." [27] After further correspondence and clarification, Einstein informed Kaluza that he submitted his paper to the academy for publication and he added: "Your idea is truly tempting. There must be some truth to it somewhere."[28]

Einstein continued to think actively about the five-dimensional spacetime approach to the unification goal. He worked on it with his assistant Jakob Grommer (see box). They explored the possibility of spherically symmetric solutions of the field equations in Kaluza's theory, which could be interpreted as representing material particles, and they formulated a criterion for the existence of such solution. Their conclusion was that "Kaluza's theory possesses no centrally symmetric solution dependent on the $g_{\mu\nu}$'s alone that could be interpreted as a (singularity-free) electron."[29] This topic would recur again and again in Einstein's later work on unified field theories.

THE HEBREW UNIVERSITY CONNECTION

The paper by Einstein and Grommer, "Proof of the Non-Existence of Everywhere Regular, Centrally Symmetric Field According to the Field Theory of Kaluza," was published in a collection of papers in physics and mathematics entitled *Scripta Universitatis atque Bibliothecae Hierosolymitanarum, Mathematica et Physica* 1 (1923). This volume appeared two years before the official opening of the Hebrew University of Jerusalem and is its first scientific publication. Every article in this volume was printed in the language of the authors and translated into Hebrew. Grommer himself translated this article. Einstein was at that time a member of the board of governors of the Hebrew University-to-be and also chairman of its Academic Committee. As a gesture to the university, he gave his name as editor of this volume.

Einstein's interest in the five-dimensional approach resulted in two papers on Kaluza's theory in 1927. At about the same time, this idea was picked up by the physicist Oskar Klein. His hope was that a unified field theory could also account for quantum phenomena. In the context of the unification scheme, Klein's main idea was that the additional

dimension was wound up in very small cylindrical surface, too small to be detected. This notion of the Kaluza-Klein theory resurfaced in string theory in the form of the additional compact dimensions (see box).

STRING THEORY

Einstein's unification program failed to achieve its goals. But the quest for unification of all the forces in nature is the goal of modern string theory as well as of other approaches. String theory is a theoretical framework in which the point-like particles are replaced by one-dimensional objects called strings. It is an attempt to explain all of the particles and fundamental forces of nature in one theory. Consistency of the theory requires the addition of six space dimensions to the four-dimensional spacetime. The additional dimensions are compactified, just like the additional space dimension in the Kaluza-Klein theory is compactified in a circle and describes electromagnetism from the perspective of a large space of three dimensions. Thus, the original Kaluza-Klein theory is a precursor to the unification of gravity and the other interactions currently pursued by string theory.

C. EINSTEIN'S UNIFICATION PROGRAM—THE NON-SYMMETRIC FIELD

In 1925, Einstein published his first original attempt to unify the gravitational and electromagnetic fields. The metric tensor and the affine connection were treated on the same footing as the fundamental starting points: both were assumed to be asymmetric. Einstein failed in his attempt to carry this program to a physically sound theory. He abandoned it and for twenty years explored other ways to achieve the goal of unification. His attempts have been investigated by the historians Jeroen van Dongen and Tilman Sauer.[30] From this time on, Einstein saw the program as an alternative to quantum mechanics, which he did not believe would be able to yield a satisfactory solution to the quantum problem, mainly because of its probabilistic nature. He erroneously believed that just pursuing mathematical formulations with a quest for simplicity had led him earlier to his successful formulation of general relativity. As van Dongen has argued, based on this self-deception, he continually tried new mathematical approaches and then discarded them when they did not produce the expected results.

Around 1945, Einstein returned to the approach he had tried twenty years earlier. Again, it was based on a non-symmetric metric tensor in which all the sixteen elements are independent functions of spacetime coordinates. Ten of them, or rather ten combinations of them, would represent the gravitational field as in the theory of general relativity, and the other six were expected to represent electromagnetism, which is characterized by the six components of the electric and magnetic vector fields. Einstein pursued this approach during the last ten years of his life.

To his longtime colleague Erwin Schrödinger, Einstein sent his first attempts along these lines:

> I am sending you herewith two papers. . . . I am sending them to nobody else, because you are the only person known to me who is not wearing blinders in regard

> to the fundamental questions in our science. The attempt depends on an idea that at first seems antiquated and unprofitable, the introduction of a non-symmetric tensor g_{ik} as the only physically relevant field quantities. . . . Pauli sticks his tongue at me whenever I tell him about it.[31]

Pauli's objection was based on the fact that the symmetric and anti-symmetric parts are both tensors and transform independently so that their sum is not a unified structure. Einstein did not think that the theory should be rejected on the basis of this argument. He asked Schrödinger's opinion on this.

Schrödinger responded with two detailed letters posing questions and raising critical remarks about Einstein's theory. Einstein was very grateful and encouraged by Schrödinger's interest in his work:

> I am very enthusiastic that you refer to my new hobby and this in such a short amount of time. It is really admirable. If I would be a fairly good person, I would thank you and leave you alone. But I am not capable of doing so and must answer your remarks.[32]

These remarks and responses evolved during the following months into an intensive correspondence with more than a dozen letters exchanged between the two colleagues.

Struggling with the difficulties of his undertaking and frustrated with lack of tangible progress Einstein wrote to Schrödinger:

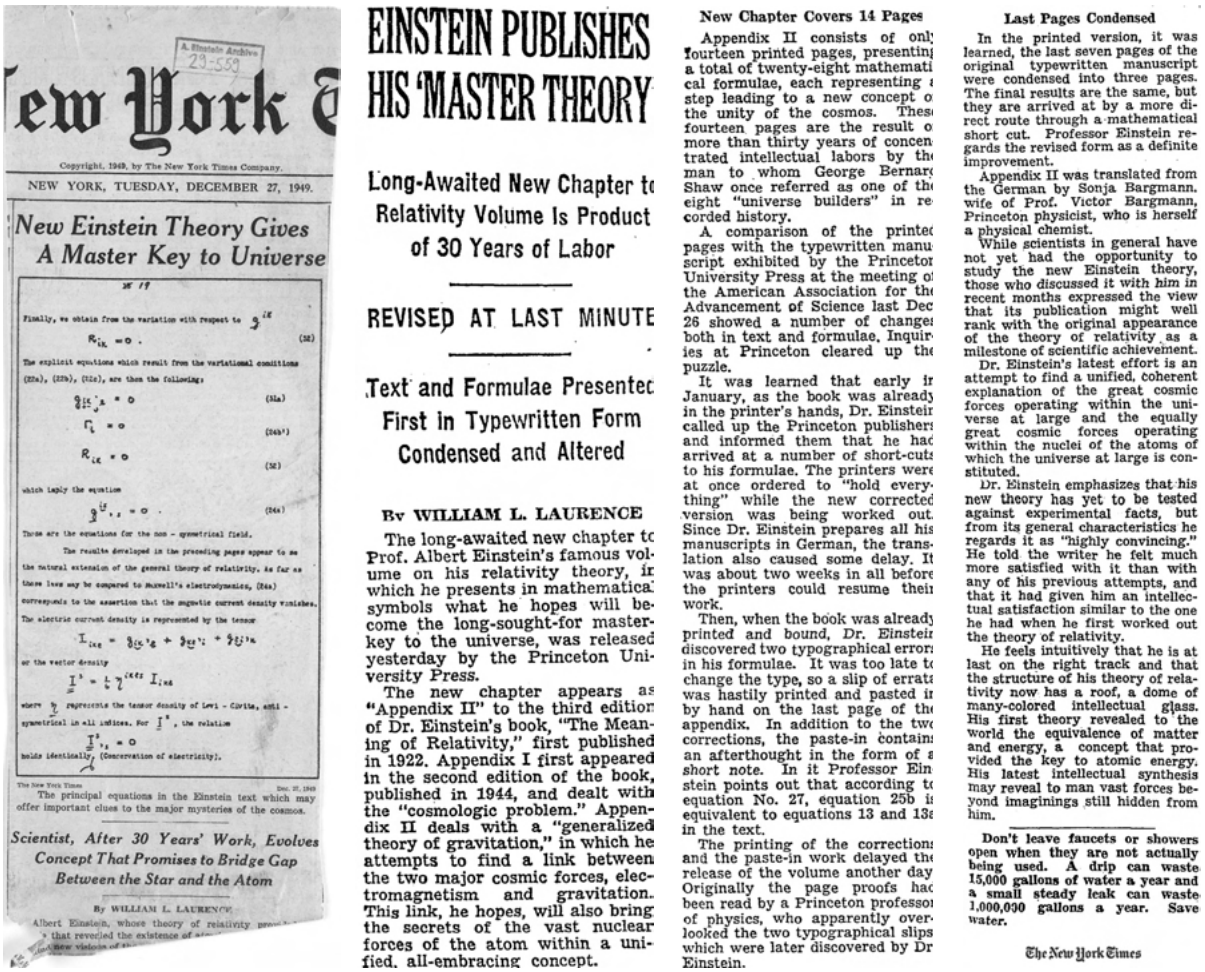
> How much I do understand your hesitating attitude. I must tell you right away that deep inside I am not so certain as I previously asserted. . . . We have squandered a lot of time on this and the result looks like a gift from the devil's grandmother.[33]

In spite of his uncertainty, Einstein continued to work on this strategy, producing paper after paper and generating sensational headlines in the *New York Times* (see figure 18). He summarized his first years of work along these lines and published it as Appendix II in the 1950 edition of *The Meaning of Relativity*. Yet his frustration and pessimism continued. He shared his feelings with his old friend Maurice Solovine:

> You might imagine that I look back on my life's work with silent content. But it is totally different when looked upon from a personal point of view. There is not a single concept about which I am convinced that it will last and I am uncertain as to whether I was even on the right track. The current generation sees in me both a heretic and a reactionary who has so to speak outlived himself.[34]

About two years later, he conveyed to Solovine the same mood, referring specifically to his search for a unified field theory:

> The unified field theory has been put into retirement. It is so difficult to employ mathematically that I have not been able to verify it. This state of affairs will last for many more years, mainly because physicists have no understanding of logical and philosophical arguments.[35]



ew York T

Copyright, 1949, by The New York Times Company.

NEW YORK, TUESDAY, DECEMBER 27, 1949.

New Einstein Theory Gives A Master Key to Universe

Scientist, After 30 Years' Work, Evolves Concept That Promises to Bridge Gap Between the Star and the Atom

By WILLIAM L. LAURENCE

EINSTEIN PUBLISHES HIS 'MASTER THEORY'

Long-Awaited New Chapter to Relativity Volume Is Product of 30 Years of Labor

REVISED AT LAST MINUTE

Text and Formulae Presented First in Typewritten Form Condensed and Altered

By WILLIAM L. LAURENCE

The long-awaited new chapter to Prof. Albert Einstein's famous volume on his relativity theory, in which he presents in mathematical symbols what he hopes will become the long-sought-for master-key to the universe, was released yesterday by the Princeton University Press.

The new chapter appears as "Appendix II" to the third edition of Dr. Einstein's book, "The Meaning of Relativity," first published in 1922. Appendix I first appeared in the second edition of the book, published in 1944, and dealt with the "cosmologic problem." Appendix II deals with a "generalized theory of gravitation," in which he attempts to find a link between the two major cosmic forces, electromagnetism and gravitation. This link, he hopes, will also bring the secrets of the vast nuclear forces of the atom within a unified, all-embracing concept.

New Chapter Covers 14 Pages

Appendix II consists of only fourteen printed pages, presenting a total of twenty-eight mathematical formulae, each representing a step leading to a new concept of the unity of the cosmos. These fourteen pages are the result of more than thirty years of concentrated intellectual labors by the man to whom George Bernard Shaw once referred as one of the eight "universe builders" in recorded history.

A comparison of the printed pages with the typewritten manuscript exhibited by the Princeton University Press at the meeting of the American Association for the Advancement of Science last Dec. 26 showed a number of changes both in text and formulae. Inquiries at Princeton cleared up the puzzle.

It was learned that early in January, as the book was already in the printer's hands, Dr. Einstein called up the Princeton publishers and informed them that he had arrived at a number of short-cuts to his formulae. The printers were at once ordered to "hold everything" while the new corrected version was being worked out. Since Dr. Einstein prepares all his manuscripts in German, the translation also caused some delay. It was about two weeks in all before the printers could resume their work.

Then, when the book was already printed and bound, Dr. Einstein discovered two typographical errors in his formulae. It was too late to change the type, so a slip of errata was hastily printed and pasted in by hand on the last page of the appendix. In addition to the two corrections, the paste-in contains an afterthought in the form of a short note. In it Professor Einstein points out that according to equation No. 27, equation 25b is equivalent to equations 13 and 13a in the text.

The printing of the corrections and the paste-in work delayed the release of the volume another day. Originally the page proofs had been read by a Princeton professor of physics, who apparently overlooked the two typographical slips which were later discovered by Dr. Einstein.

Last Pages Condensed

In the printed version, it was learned, the last seven pages of the original typewritten manuscript were condensed into three pages. The final results are the same, but they are arrived at by a more direct route through a mathematical short cut. Professor Einstein regards the revised form as a definite improvement.

Appendix II was translated from the German by Sonja Bargmann, wife of Prof. Victor Bargmann, Princeton physicist, who is herself a physical chemist.

While scientists in general have not yet had the opportunity to study the new Einstein theory, those who discussed it with him in recent months expressed the view that its publication might well rank with the original appearance of the theory of relativity as a milestone of scientific achievement.

Dr. Einstein's latest effort is an attempt to find a unified, coherent explanation of the great cosmic forces operating within the universe at large and the equally great cosmic forces operating within the nuclei of the atoms of which the universe at large is constituted.

Dr. Einstein emphasizes that his new theory has yet to be tested against experimental facts, but from its general characteristics he regards it as "highly convincing." He told the writer he felt much more satisfied with it than with any of his previous attempts, and that it had given him an intellectual satisfaction similar to the one he had when he first worked out the theory of relativity.

He feels intuitively that he is at last on the right track and that the structure of his theory of relativity now has a roof, a dome of many-colored intellectual glass. His first theory revealed to the world the equivalence of matter and energy, a concept that provided the key to atomic energy. His latest intellectual synthesis may reveal to man vast forces beyond imaginings still hidden from him.

Don't leave faucets or showers open when they are not actually being used. A drip can waste 15,000 gallons of water a year and a small steady leak can waste 1,000,000 gallons a year. Save water.

The New York Times

FIGURE 18: Sensational headlines. From the *New York Times,* February 15, 1950. Copyright 1950 by the *New York Times;* from the *New York Times,* December 27, 1949. Copyright 1949 by the *New York Times.* All rights reserved. Used by permission.

Einstein's prophecy was correct and "this state of affairs" continues until today. Still, he did not put the unified field theory into retirement. In spite of the pessimistic mood reflected in these letters, he continued to work on this problem with zeal and perseverance, struggling with it, literally, until the end of his life. A few days before Einstein passed away in April 1955, he asked his secretary to bring his notepad, filled with calculations, to the hospital so that he could continue to work. The last page of this final notepad is shown in figure 19. Einstein never managed to complete the task he set for himself. But as we can see, he never stopped trying, up to the very last of his days.

Although Einstein did not succeed in achieving his goal, the quest for a comprehensive field theory unifying all the forces in nature, a "theory of everything," is still considered by many as the greatest challenge at the frontier of theoretical physics.

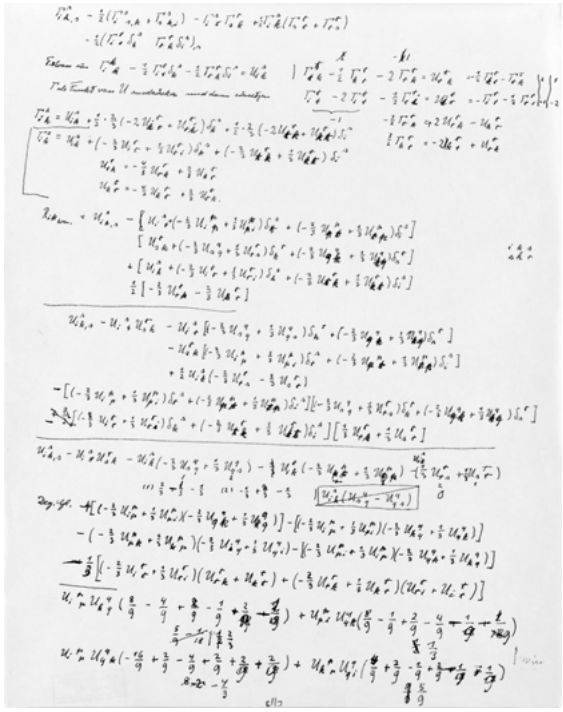


FIGURE 19: The last page of Einstein's last notepad, which he worked on until he passed away in April 1955 (AEA 3010).

D. *THE MEANING OF RELATIVITY*—APPENDIX II

Einstein presented a summary of his non-symmetric field approach in Appendix II, "Generalized Theory of Gravitation," which was added to the third Princeton University Press (PUP) edition of *The Meaning of Relativity* in 1950. He significantly expanded his presentation and published it as Appendix II, under the same title, in the 1951 Methuen edition. A modified version of this appendix was added to the fourth PUP edition in 1953 under the title "Generalization of Gravitation Theory." Einstein significantly reformed this appendix for the fifth PUP edition (see box), which appeared after his death in 1956.

A NOTE ON THE FIFTH EDITION

For the present edition I have completely revised the "Generalization of Gravitation Theory" under the title "Relativistic Theory of the Non-symmetric Field." For I have succeeded—in part in collaboration with my assistant B. Kaufman—in simplifying the derivations as well as the form of the field equations. The whole theory becomes thereby more transparent, without changing its contents.

A. EINSTEIN, DECEMBER 1954

The contents of the different versions of Appendix II represent Einstein's work long after the formative years on which our book concentrates. Nevertheless, we find it appropriate to discuss it briefly here because, through Einstein's wish, they became connected with *The Meaning of Relativity.* The different versions of this appendix summarize his persevering struggle to meet the challenge of finding a unified mathematical representation of the entire physical reality. The non-symmetric field seemed to him, at this stage of his life, the most natural approach in achieving this goal.

Einstein first points out the provisional nature of the gravitational field equation, which is reflected in the dualistic treatment of the problem when the source of the field is represented by a tensor, which itself is of nongravitational character. The theory is based on the equivalence principle, which implies the unity of gravitation and inertia. But this principle cannot serve as a guideline for formulating a more general theory:

> The principle of equivalence, however, does not give any clue as to what may be the more comprehensive mathematical structure on which to base the treatment of the total field comprising the entire physical reality. (p. 293 [133])

The lack of such a guiding principle leads to an apparently vague formulation of the problem of extending the equation of pure gravitation. Nevertheless, Einstein believed that he had found a convincing solution to this problem although he could not substantiate his results with experimental evidence.

The mathematical formalism is extremely complicated and we shall not attempt to describe it in any detail. Einstein applied the tensor calculation that he performed in the derivation of his general theory of relativity to the non-symmetrical metric to generate the analogues of the Riemann and Ricci tensors and then proceeded to derive the field equations from a Hamiltonian principle. The result seemed to Einstein to represent a natural extension of the general theory of relativity, because certain equations, which appeared in the course of this calculation, could be interpreted in the context of Maxwell's electrodynamics. One of them corresponded to the vanishing of the magnetic current and another one to the conservation of the electrical charge-current vector.

The mathematical formulation becomes more and more elaborate from one version of the appendix to the other. It is accompanied by epistemological remarks about the "strength" and nature of physical theories, a discussion of the structure of the theory of general relativity, and an illuminating explanation of Einstein's motivation to make such a laborious effort in achieving the goal of unification. In this book, we reproduce the versions of this appendix published in the third and fourth PUP editions. These versions contain extensive explanatory background texts, full of interesting ideas, which do not appear in the previous versions, and the mathematical formalism (for those who are interested and able to follow it) is explained in greater detail.

The first of the two versions included here contains a section entitled "General Remarks Concerning the Concepts and Methods of Theoretical Physics" (p. 321 [161]). It is a self-contained epistemological essay on the basic elements of contemporary physical theories, such as the spacetime continuum and the essential role of the objective, independent of representation, existence of a point. To a point in one representation there corresponds a point in another representation of the physical system, namely, only one-valued coordinate transformations are allowed. Einstein emphasizes this element because

it implies the indispensable role of tensors and vectors in present theories. Then he adds wild speculations of possible theories of an as yet unexplored nature:

> A priori, though, one cannot see why there could not be equivalent representations of a physical situation in which the identity of the point is not preserved (while the continuum is maintained). I have, however no idea how to arrive at a relativistic theory in this wider sense. (p. 324 [164])

However, the concluding remarks in this appendix deserve special attention and explicit exposure. Let us quote from them at length:

> I must, however, explain why I have gone to so much trouble to arrive at this result. The contemporary physicist cannot, without such an explanation, appreciate this; for he is convinced, as a result of the successes of the probability-based quantum mechanics, that one must abandon the goal of complete descriptions of real situations in a physical theory. I do not want to discuss here why I do not share this conviction . . .
>
> There is further the conviction that one cannot keep side by side the concepts of fields and particles as elements of the physical description. . . . The field concept, however, seems to be inevitable, since it would be impossible to formulate general relativity without it. . . . For this reason I see in the present situation no possible way other than a pure field theory, which then however has before it the gigantic task of deriving the atomic character of energy. (p. 324–325 [164–165])

These remarks confirm that Einstein had undertaken this painstaking effort to formulate a comprehensive unified field theory in the hope that it could be an alternative to the contemporary probabilistic interpretation of quantum mechanics, which, for him, was an unacceptable description of physical reality.

He returns to this point in the concluding remarks of the later version of the appendix:

> Is it conceivable that a field theory permits one to understand the atomistic and quantum structure of reality? Almost everybody will answer this question with "no." But I believe that at the present time nobody knows anything reliable about it.

Einstein refers here to a classical nonlinear field theory, like the general theory of relativity, or its extension to a unifying field theory. In his time, mathematical difficulties prevented him from drawing definite conclusions from such theories about the existence and effect of singularity-free solutions for such theories. In his view, this situation precluded the possibility of "comparing the content of non-linear field theories with experience." As for the quantum field theories adopted and explored by the mainstream physics community, Einstein's attitude was that

> at the present time the opinion prevails that a field theory must first, by "quantization", be transformed into a statistical theory of field probabilities according to more or less established rules. I see in this method only an attempt to describe relationships of an essentially nonlinear character by linear methods.

In the concluding remarks of this version of the appendix, as in the previous version, we find another speculative idea, that of providing a purely algebraic description of reality:

> One can give good reasons why reality cannot at all be represented by a continuous field. From the quantum phenomena it appears to follow with certainty that a finite system of finite energy can be completely described by a finite set of numbers (quantum numbers). This does not seem to be in accordance with a continuum theory, and must lead to an attempt to find a purely algebraic theory for the description of reality. But nobody knows how to obtain the basis of such a theory.[36]

NOTES

1. For detailed surveys of different approaches adopted by Einstein toward this goal, see Jeroen van Dongen, *Einstein's Unification* (Cambridge: Cambridge University Press, 2010); and Tilman Sauer, "Einstein's Unified Field Theory Program," in *The Cambridge Companion to Einstein*, pp. 281–305.
2. Albert Einstein, "Fundamental Ideas and Problems of the Theory of Relativity," in *Les Prix Nobel en 1921–1922*, ed. C. G. Santesson (Stockholm: Nobel Foundation, 1923), p. 489, http://www.nobelprize.org/nobel_prizes/physics/laureates/1921/einstein-lecture.pdf.
3. Albert Einstein, "On the General Theory of Relativity," CPAE vol. 6, Doc. 21.
4. Albert Einstein, "On the General Theory of Relativity (Addendum)," CPAE vol. 6, Doc. 22, p. 108.
5. Albert Einstein, "The Foundation of General Relativity," CPAE vol. 6, Doc. 30, p. 188.
6. Albert Einstein, "Do Gravitational Fields Play an Essential Role in the Structure of the Elementary Particles of Matter," CPAE vol. 7, Doc 17.
7. CPAE vol. 7, Doc 17, p. 88.
8. The fourth Princeton lecture, p. 258 [98].
9. "Einstein Cannot Measure the Universe: With Mean Density of Matter Unknown the Problem is Impossible. Final Princeton Lecture," *New York Times*, 14 May 1921.
10. Hermann Weyl to Einstein, 1 March 1918, CPAE vol. 8, Doc. 472.
11. Einstein to Hermann Weyl, 6 April 1918, CPAE vol. 8, Doc. 498.
12. Einstein to Hermann Weyl, 8 April 1918, CPAE vol. 8, Doc. 499.
13. Reprinted in CPAE, German version, vol. 7, Doc. 8. Authors' translation.
14. Einstein to Heinrich Zangger, 1 June 1919, CPAE vol. 9, Doc. 52.
15. Einstein to Arthur S. Eddington, 15 December 1919, CPAE vol. 9, Doc. 216.
16. Einstein to Paul Ehrenfest, 2 February 1920, CPAE vol. 9, Doc. 294.
17. Einstein to Heinrich Zangger, 27 February 1920, CPAE vol. 9, Doc. 332.
18. Eddington, *Space, Time and Gravitation*, p. 167.
19. Einstein, in the introduction to Mario Pantaleo, *Cinquant'Anni di Relatività* (Florence: Editrice universitaria, 1955). Translated by John Stachel.
20. "On the General Theory of Relativity," CPAE vol. 13, Doc. 425, p. 371.
21. Einstein to Niels Bohr, 10 January 1923, CPAE vol. 13, Doc. 421.
22. Albert Einstein, "The Theory of the Affine Field," *Nature* 112 (23 September 1923): 448–449; reprinted in CPAE vol. 14 (German edition), Doc. 123.
23. Einstein to Theodor Kaluza, 21 April 1919, CPAE vol. 9. Doc. 26.
24. Ibid.
25. Einstein to Theodor Kaluza, 28 April 1919, CPAE vol. 9, Doc. 30.
26. Einstein to Theodor Kaluza, 14 May 1919, CPAE vol. 9, Doc. 40.
27. Einstein to Theodor Kaluza, 14 October 1921, CPAE vol. 12, Doc. 270.
28. Einstein to Theodor Kaluza, 9 December 1921, CPAE vol. 12, Doc. 318.
29. "Proof of the Non-Existence of an Everywhere Regular Centrally Symmetric Field, According to the Field Theory of Kaluza," CPAE vol. 13, Doc. 12, p. 33.

30. See Jeroen van Dongen, *Einstein's Unification* (Cambridge: Cambridge University Press, 2010); and Sauer, "Einstein's Unified Field Theory Program."
31. Einstein to Erwin Schrödinger, 22 January 1946, AEA 22-93.
32. Erwin Schrödinger to Einstein, 22 February 1946, AEA 22-98.
33. Einstein to Erwin Schrödinger, 20 May 1946, AEA 22-106.
34. Einstein to Maurice Solovine, 28 March 1949, AEA 21-260.
35. Einstein to Maurice Solovine, 12 February 1951, AEA 21-277.
36. See John Stachel, "The Other Einstein: Einstein Contra Field Theory," *Science in Context* 6:1 (1993): 275–290.

9
EARLY MONOGRAPHS ON RELATIVITY

In the previous chapters, we discussed various aspects of the process of revision and refinement of the basic ideas underlying the general theory of relativity and the controversial debates about their consequences during the formative period of Einstein's new theory. This process progressed through articles published in scientific journals and through extensive correspondence between the physicists, mathematicians, astronomers, and philosophers who formed a network of scientists interested in these developments. This process was also accompanied by public and university lectures and by the publication of a series of monographs and textbooks that contributed to the dissemination and acceptance of Einstein's theory in the scientific community worldwide, to its teaching in universities, and also to its comprehensible presentation for interested but nonprofessional readers.

Most of the books that are considered in this chapter appeared in the years 1920–1921. This sudden and broad interest in the theory of relativity, and in Einstein himself, was triggered by the confirmation, in November 1919, of the prediction of the theory that a beam of light would be bent by the gravitational field. Let us quote from Sommerfeld's introduction to Pauli's book, to be discussed below, in order to illustrate this point:

> In view of the apparently insatiable demand, especially in Germany, for accounts of the Theory of Relativity, both of a popular and of a highly specialized kind, I felt I ought to advise the publishers to arrange for a separate edition of the excellent article by Mr. W. Pauli, jr., which appeared in the *Encyclopädie der mathematischen Wissenschaften*, Vol. V.[1]

Although the appeal of sensationalism may have been part of the reason for this outburst of interest, which turned Einstein into a world celebrity almost overnight, there was certainly also a genuine quest for understanding these developments.

The innovative and counterintuitive nature of Einstein's new theory also provoked a flood of popular discussions and publications, ranging from newspaper articles to popular books. They comprise critiques of different categories, ranging from genuine misunderstandings to vicious anti-Semitic polemics. This phenomenon will be discussed toward the end of this chapter.

A. THE FIRST MONOGRAPHS—FREUNDLICH AND EINSTEIN

The Foundations of Einstein's Theory of Gravitation, by Erwin Freundlich, was the first brief account of this subject. It presented the theory without the mathematical formalism and emphasized the basic ideas, which led from the special to the general theory of relativity and to the gravitational field equation. It focused on the experimental tests of the validity of the new theory in connection with Freundlich's own work. At the time of the publication of this book, the experimental confirmation of Einstein's theory was still inconclusive and constituted a topic of controversial discussion. Freundlich summarized his treatise:

> In spite of this, the theory can, even at this early stage, justly claim general attention, the reason is to be found in the unusual unity and logical structure of the ideas underlying it. In truth it solves, at one stroke, all the riddles, concerning the motion of bodies, which have presented themselves since the time of Newton, as the result of the conventional view about the meaning of space and time in the physical description of natural phenomena.[2]

At that time, Einstein's attitude to Freundlich was marked by appreciation and gratitude, which is reflected in his complimentary introduction to Freundlich's book:

> Dr. Freundlich has undertaken in the following essay to illumine the ideas and observations which gave rise to the general theory of relativity so as to make them available to a wider circle of readers. I have gained the impression in perusing these pages that the author has succeeded in rendering the fundamental ideas of the theory accessible to all who are to some extent conversant with the methods of reasoning of the exact sciences. The relations of the problem to mathematics, to the theory of knowledge, physics and astronomy are expounded in a fascinating style, and the depth of thought of Riemann, a mathematician so far in advance of his time, has in particular received warm appreciation. Dr. Freundlich is not only highly qualified as a specialist in the various branches of knowledge involved to demonstrate the subject; he is also the first amongst fellow-scientists who has taken pains to put the theory to the test. May his booklet prove a source of pleasure to many.[3]

However, before the publication of the third edition, Einstein wrote to Freundlich telling him that the book needed some improvement and that he would withdraw his preface if appropriate corrections were not made in the forthcoming edition:

> First, as regards your booklet, quite some correction is in fact necessary if gross misunderstandings are to be avoided. I shall be glad to talk everything over with you . . . If we are not able to agree on the points . . . my brief foreword endorsing it would just have to be omitted.[4]

They did agree on the necessary corrections, and in the introduction to the third edition, Freundlich thanked Einstein for his help in the revised version of the booklet.

The second monograph exposing the ideas of the theory of relativity was Einstein's own. When he was summarizing the final version of his general theory of relativity, Einstein

was already thinking about writing a popular book on the subject. In January 1916 he wrote to his close friend Michele Besso:

> The great success in gravitation pleases me immensely. I am seriously contemplating writing a book in the near future on special and general relativity theory, although, as with all things that are not supported by a fervent wish, I am having difficulty getting started. But if I do not do so, the theory will not be understood, as simple though it basically is.[5]

Einstein completed the manuscript in December, and the booklet (as he referred to it) *Relativity: The Special and the General Theory (A Popular Account)* was published in the spring of 1917.[6] It has already been mentioned in the introduction to this book that it may be viewed as a nonprofessional precursor to *The Meaning of Relativity.* As stated in the preface, Einstein "spared himself no pains in his endeavor to present the main ideas in the simplest and most intelligible form."[7] Yet it is not a popular book in the usual sense. It may be popular in its format, in the dialogues with the reader, in its examples from daily life, and in the lack of mathematical formulas, but it does not compromise on scientific rigor, and the reader soon discovers that an intellectual effort is required to follow the flow of Einstein's thoughts and arguments. But Einstein was not happy with the result. In a letter to Besso he wrote:

> The description has turned out quite wooden. In the future, I shall leave writing to someone else whose speech comes more easily than mine and whose body is more in order.[8]

Yet this booklet is an attempt to enable the general reader, with no background in physics, to grasp and appreciate the grandeur of one of the most sophisticated intellectual achievements of the human mind and thereby to grant him, as Einstein puts it, "a few happy hours of suggestive thought."[9] Einstein appeals to the reader's intuition, not assuming any previous knowledge of the subject matter. He uses familiar metaphors like train carriages and embankments. He often invites the reader to participate in the thought process by posing a question to the reader, which he answers himself or in the name of the reader. Such parts of the text resemble Galileo's famous dialogs, and invite the reader to actively participate in the thought process.

Einstein's booklet is written with sophistication and elegance. Ideas that in hindsight appear as erroneous, such as the conception of a mechanical ether as a carrier of light waves and as a unique frame of reference, are not simply discarded as deviations from the right path but are instead regarded as intellectual challenges that had to be addressed before they could be replaced by the ideas of modern science. Einstein takes the reader on a smooth path from classical mechanics, through the special theory of relativity, to the general theory and its immediate consequences. There is hardly any trace of the bumpy road and of the difficulties he encountered on his way to this achievement.

This book was hugely successful. Between 1917 and 1922 it appeared in fourteen editions in German. During his lifetime there were fifteen editions in German. After the confirmation of the bending of light, it also appeared in many foreign languages. The success of Einstein's booklet in Germany was followed by a similar, unprecedented

success in other countries. The first English translation was published in 1920, followed by the French (1921), Italian (1921), Japanese (1921), Polish (1921), Russian (1921), Spanish (1921), Hungarian (1921), Chinese (1922), Czech (1923), and Hebrew (1928) editions. Some of these foreign-language editions also saw, within a short period, new editions. They played a significant role in the spread of the theory of relativity and its public understanding in the respective countries. To the third edition of his book, Einstein added a remark:

> This year (1918) Springer-Verlag published an elaborate excellent textbook on the general theory of relativity, written be H. Weyl with the title *Space, Time, Matter*. It is to be warmly recommended to mathematicians and physicists.

B. HERMANN WEYL: GENERAL RELATIVITY REFORMULATED AND EXTENDED

After Einstein's *The Foundations of General Relativity*, Weyl's book *Space, Time, Matter* (1918) was the first comprehensive account of the general theory of relativity with its mathematical formalism. It was quoted more than the other books discussed in this chapter and was mentioned extensively in almost all of them. It had a major impact on the formulation of the relativistic theory of gravitation in its formative years and on the debates of its consequences, refining the mathematical tools and combining them with the physical contents of the theory.

This book was based on lectures delivered in the summer of 1917 at the Eidgenössische Technische Hochschule (ETH) in Zurich. Weyl had been motivated to publish these lectures in book form because, as stated in the introduction, a systematic presentation of the theory was lacking. It was his desire to present this subject as an interwoven process of philosophical, mathematical, and physical thinking, building the theory systematically from its foundations with special attention to the principles. He did not think that he succeeded in this endeavor, because "the mathematician predominates at the expense of the philosopher."[10]

Einstein responded to Weyl's book with great appreciation as soon as he had read the printer's page proofs,[11] admiring the mathematical ingenuity and clarity of the presentation. When the book appeared in print, he praised it again in a review article that he published. It is worth quoting from this review at length:

> I am always tempted to read the individual parts of this book again, because every page shows the amazingly steady hand of the master who has penetrated the subject matter from the most diverse angles. I consider it a happy occasion that such a distinguished mathematician has taken care of this new field. He understood how to combine mathematical rigor with graphic intuition. From this book, the physicist can learn the foundations of geometry and the theory of invariants, and the mathematician can learn those of electricity and the theory of gravitation. . . .
>
> The second chapter is an introduction to the absolute differential calculus and Riemannian geometry, resp. One especially sees there with amazement how the most complicated becomes simple and self-evident under Weyl's hand. . . .

> Recently, Weyl has earned considerable merits by the integration of the field equations of gravitation. The expositions of the last paragraphs exemplify how a *born* mathematician can be effective here through simplifying and clarifying.

In spite of a few critical remarks, Einstein concluded that

> the effort to study it is brilliantly rewarded, and there is perhaps hardly anyone who could not learn from this book.[12]

Weyl's book appeared in five editions until 1922. Of particular significance is the third edition, which appeared in 1919. As we discussed in chapter 8, the introduction of the concepts of affine connection and infinitesimal parallel displacements by Levi-Civita motivated Weyl to reexamine the mathematical formulation of Einstein's theory of gravitation from the perspective of a reformulation of Riemannian geometry and its application to the derivation of the gravitational field equation. This is the formalism that Einstein used in *The Meaning of Relativity*. Weyl also discussed in this edition his extension of Riemannian geometry to include parallel displacements, which do not preserve the length of a vector. This was the basis of Weyl's model of a field theory, which unifies Einstein's gravitational field equation with Maxwell's equation of electromagnetism.[13] As mentioned above, in letters to colleagues (chapter 8, p. 127 ff.), Einstein was critical of the third edition of this book because of the inclusion of this model of unification and also because of Weyl's treatment of energy-momentum conservation.

C. ARTHUR EDDINGTON—THE FIRST BOOK IN A LANGUAGE OTHER THAN GERMAN

Arthur Eddington is best known in the context of the general theory of relativity for leading the astronomical expedition that confirmed the bending of light by the gravitational field of the sun. In addition, during the formative years, next to de Sitter he was the most prominent protagonist of general relativity outside of Germany and the author of the first comprehensive account of this subject in a language other than German.

Eddington's *Space, Time and Gravitation*, published in 1920, presents the basic ideas of the theory of relativity as a revolution of thought in physics. It traces the road from the special to the general theory of relativity and explores its cosmological consequences and the quest for generalization of Einstein's theory of gravitation to include electromagnetism. Except for a few mathematical notes at the end and several mathematical expressions scattered throughout the text, the book avoids mathematical formalism. The author states that his aim is to give an account "without introducing anything technical in the way of mathematics, physics and philosophy."[14] Although not expressed in the symbolic language of mathematics, the underlying mathematical concepts and methods are presented and explained. The text is also permeated with epistemological and philosophical discussions. Thus, the prologue, "What Is Geometry?," is a conversation between an experimental physicist, a mathematician, and a relativist, who advocate the new conceptions of time and space in physics. The last chapter, "On the Nature of Things," is more philosophical. It describes Eddington's views on the relation between mathematical concepts associated with the theory of relativity, such as four-dimensional spacetime, Riemannian geometry, and more general geometries on the one hand, and

physical reality on the other. Before this chapter, Eddington devotes a whole chapter to the discussion of Weyl's proposal to unify electricity and magnetism (chapter 8, p. 129), indicating that this approach may lead to an understanding of the nature of matter, specifically of the stability of charged particles.

Eddington's second book, *The Mathematical Theory of Relativity* (1923), is, to a large extent, based on the mathematical supplement to the French edition of his *Space, Time and Gravitation.* The text has been modified and expanded to about three times the size of its original version. In the introductory remarks, the author expresses his hope that this more systematic and comprehensive treatise of the mathematical formalism of relativity "may meet the needs of those who wish to enter fully into these problems of reconstruction of theoretical physics." Referring to his previous book, he justifies its nontechnical nature, saying:

> It is eminently desirable to have a general grasp of the revolution of thought associated with the theory of Relativity before approaching it along the narrow lines of strict mathematical deduction. In the former work we explained how the older conceptions of physics have become untenable, and traced the gradual ascent to the ideas which must supplant them.[15]

The author concludes that the theory of relativity "appears to cover fairly adequately the phenomena of physics" except for an explanation of electrons or quanta. The book ends with the question if the theory of relativity will ultimately be able to account also for the phenomena of the microscopic atomistic world of physics.

> On the one hand it would seem an idle exaggeration to claim that the magnificent conception of Einstein is necessarily the key to all the riddles of the universe; on the other hand we have no reason to think that all the consequences of this conception have become apparent in a few short years.[16]

D. MAX VON LAUE—"A BOOK WRITTEN BY A PHYSICIST"

Max von Laue published his book on general relativity and Einstein's theory of gravitation in 1921 as *The Theory of Relativity*, the second volume. The first volume was first published in 1911 under the subtitle *The Principle of Relativity and the Lorentz Transformation* and was republished with modifications and expansions in 1913 and 1919. All the three editions were devoted to the special theory of relativity.[17]

Von Laue made important contributions to the development of relativity theory; these are described in the first volume. In particular, he explored the implications of Einstein's mass-energy relation, $E = mc^2$, for a deeper understanding of relativistic continuum dynamics. According to the mass-energy relation, stresses that are familiar from elasticity theory or hydrodynamics also embody energy and may thus change the inertial properties of a moving body. In fact, all forms of energy must have mass. In contrast, in classical physics stresses have no effect on the motion of a body. Even when dealing with an extended body, in classical physics it is always possible to describe the effect of forces on its overall motion—leaving aside deformations—by a single quantity, its inertial mass. In relativity, however, there is generally no longer a single quantity, such as

mass, that would characterize the inertial behavior of an extended physical system. Von Laue's work around 1911 made it clear that for this purpose no less than ten functions are required, which together form the components of a geometric object in spacetime called the *stress-energy*, or *energy-momentum tensor*. As early as 1912, Einstein realized that this tensor must play the crucial role of the source-term in his equation for the gravitational field.[18]

Von Laue himself, however, at first did not accept Einstein's generalized theory. In August 1913, he wrote to Schlick:

> The extraordinary, in fact inconceivable complexity of the theory confirms my rejection. Fortunately, one of its most direct consequences, the bending of light rays near the sun, can be checked already in 1914 during the solar eclipse. Then the theory will likely pass away peacefully.[19]

As we have seen (p. 54 ff.), the bending of light could not be checked in 1914, and only in 1919 could observations establish its validity.

In the introduction to the second volume, von Laue admits his initial skepticism toward Einstein's theory. With time, he became more convinced, although he still thought the theory had not yet reached the level of confirmation that the special theory of relativity had:

> The author of this book was at first rather doubtful [about the theory], and thus began rather late to work on it. But the more he tried to understand the theory, the more he sensed its intrinsic persuasive power. Nevertheless, the author still thinks that the theory has not received a certain confirmation through experience, nor an inner coherence, like the previous limited [*beschränkt* in the original] relativity theory, when writing the first volume of this book. This book is not an advertisment, but we cannot help but wish that the theory in its current version might have a persuasive power on the reader.[20]

In the introduction, von Laue explained his motivation in writing this account of the general theory of relativity at a time when the theory was still approached with conflicting attitudes—admired by some and denounced by others. Among the excellent books written on the subject, he mentioned the mathematical, significant work by Hermann Weyl (*Space, Time, Matter*) and the brilliantly written philosophical book by Ernst Cassirer (*Einstein's Theory of Relativity*),[21] but there was no book written by a physicist "in a strict scientific manner and in a certain completeness." His intention was to fill this void because he believed that only a physicist was fully able to understand and rectify the difficulties, which kept the majority of his colleagues from accepting the theory. In his words:

> Most people, though, do not know how to take a certain point of view of this theory, simply because they are not well acquainted with non-Euclidian geometry and its accompanying tensor calculus. And exactly for these people my book will be of help.[22]

E. WOLFGANG PAULI—"A MATURE AND BROADLY CONCEIVED WORK"

Pauli's book is another classic, one of the best treatises on special and general relativity. Remarkably, it was written in December 1920, when Pauli was a twenty-year-old student of physics at the University of Munich, and published in September 1921 as an essay in the *Encyclopedia of Mathematical Sciences*, vol. 5. One of the aims of this article was to provide a complete review of the literature on the theory of relativity (the special and the general). At the suggestion of Sommerfeld, who was the editor of this particular volume of the encyclopedia, this essay was also published as a separate book. Sommerfeld justified this initiative by the "apparently insatiable demand, especially in Germany, for accounts of the theory of relativity, both of a popular and of a highly specialized kind."[23] The book was reviewed by Einstein. Profoundly impressed, he wrote:

> Whoever studies this mature and broadly conceived work would not believe that the author is a man of twenty-one years of age. One does not know what to admire most: the psychological grasp of the development of ideas, the certainty of mathematical deductions, the profound physical vision, the capacity for clear systematic presentation, the knowledge of the literature, the factual completeness, or the certainty of the critique.[24]

Pauli's book covers the principles and the mathematical formulation of the special and general theories of relativity and Einstein's theory of gravitation. It also covers all the major issues discussed in the first years after its completion: the first exact solutions, the cosmological models of Einstein and de Sitter, Mie's theory of the origin of charged particles, Weyl's unified field theory, and the first attempts to apply it to the study of matter. The vast bibliography in this book shows just how much had been written on these subjects in the first few years.

The last section of the book is a critical discussion of the theories dealing with the problem of matter, namely, with the attempts to derive the atomic nature of electricity from a continuum field theory. At that time, many physicists believed that this could be achieved. As discussed in the preceding chapter, Einstein held on to this belief until the end of his life. He admitted that his hopes for a solution to the problem along these lines had not been fulfilled, but he still considered it to be an open question. For him, the "unification program" also meant a theory of the elementary particles based on classical continuous fields, which are free of singularities (chapter 8, p. 132 ff.). Pauli, on the other hand, was already skeptical in 1920 about the possibility of explaining the nature of particles of matter and the atomism of electric charges on the basis of classical concepts of continuous fields. He concludes the book by stating:

> Whatever may be one's attitude in detail towards these arguments, this much seems fairly certain: new elements which are foreign to the continuum concept of the field will have to be added to the basic structure of the theories developed so far, before one can arrive at a satisfactory solution of the problem of matter.[25]

F. HANS THIRRING: THE THEORY OF RELATIVITY IN SIMPLE LANGUAGE

Hans Thirring was an Austrian theoretical physicist known, together with the mathematician Josef Lense, for predicting in 1918 what is known as the Lense-Thirring *frame-dragging* effect of general relativity (mentioned in chapter 5 above, pp. 80–81). His book *The Ideas of Einstein's Theory* was written for nonprofessional readers, and the chain of assertions is not substantiated by mathematical formalism. In the introductory remarks to the German edition, the author emphasizes that his main goal was to provide the physical basis of the theory:

> This book treats the theme from a purely physical point of view—a further philosophical evaluation of the ideas will not be attempted. Nevertheless it is precisely this book which is intended as a basis for a philosophical discussion on the theory of relativity: here essentially everything has been said that the physicist has to say; now the philosopher may take position![26]

An English translation of his book was published in the same year. The author is more specific about the contents and nature of his work in the preface to the English edition:

> The present book, though written primarily for laymen, may also be useful to those who are versed in theory from the mathematical point of view, but who may find it convenient to supplement their knowledge of the general aspects of the subject. . . . A serious exposition of the theory will not only have to lay stress on the fact of how very revolutionary Einstein's theory is from the point of view of principle and theory, but must also indicate how very non-revolutionary it appears from the practical point of view. . . . The astronomer, therefore, with few exceptions will continue to calculate according to Newton's theory, the man of science will go on using Maxwell's equations, and little will be altered. But the mental foundations of the complete system of physics have been entirely changed. This will be elucidated by numerical examples, so as to dispel wrong and fanciful ideas of the theory on the part of the reader.[27]

The book was also translated into French by Einstein's close friend, Maurice Solovine. As a gesture to his friend, Einstein wrote a few lines of introduction to the French edition:

> The book by Thirring contains one of the best expositions of the theory of relativity addressed to the general public. The synoptic schema added at the end, which illuminates the independence of the guiding ideas of the theory and way in which it interconnects them, is particularly instructive.[28]

The "synoptic schema" that so impressed Einstein is a chart at the end of the book combining concepts, hypotheses, and physical consequences of the special and the general theory of relativity and showing the relations between them (see figure 20).

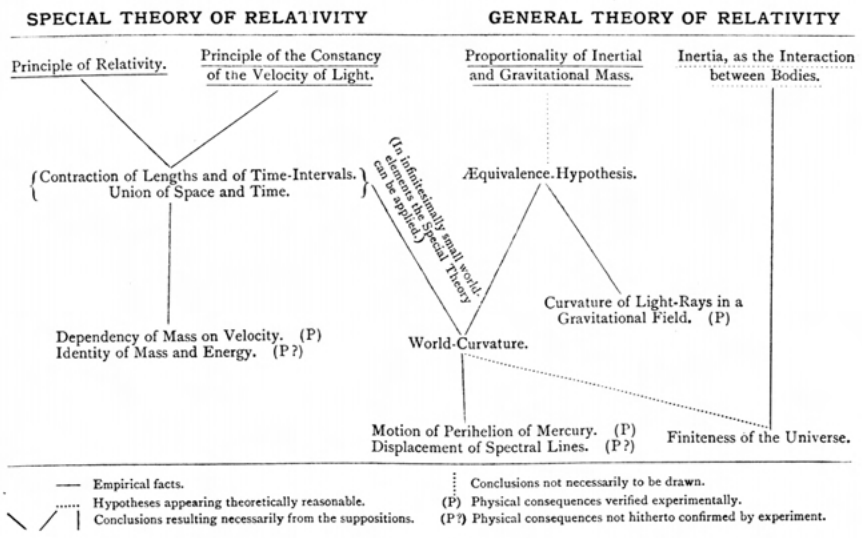

SPECIAL THEORY OF RELATIVITY
GENERAL THEORY OF RELATIVITY
Principle of Relativity.
Principle of the Constancy of the Velocity of Light.
Proportionality of Inertial and Gravitational Mass.
Inertia, as the Interaction between Bodies.
Contraction of Lengths and of Time-Intervals. Union of Space and Time.
(In infinitesimally small world-elements the Special Theory can be applied.)
Æquivalence. Hypothesis.
Curvature of Light-Rays in a Gravitational Field. (P)
Dependency of Mass on Velocity. (P)
Identity of Mass and Energy. (P?)
World-Curvature.
Motion of Perihelion of Mercury. (P)
Displacement of Spectral Lines. (P?)
Finiteness of the Universe.
Empirical facts.
Hypotheses appearing theoretically reasonable.
Conclusions resulting necessarily from the suppositions.
Conclusions not necessarily to be drawn.
(P) Physical consequences verified experimentally.
(P?) Physical consequences not hitherto confirmed by experiment.

FIGURE 20: "Synoptic schema," from Hans Thirring, *The Ideas of Einstein's Theory*, p. 167.

G. FROM LECTURE SERIES TO MONOGRAPHS

The formative years also saw the first courses on general relativity at universities, as well as public lecture series that responded to the broad interest in this subject. A number of books were written on the basis of such lecture series. We mention here three examples.

Max Born. *Einstein's Theory of Relativity*. This book is the printed version of a series of lectures delivered to a large nonprofessional audience in Frankfurt in 1920. As stated in the introduction, this was a time of growing interest in the theory of relativity and in Einstein as both a scientist and a human being. Despite a genuine desire to understand the theory, the level of mathematical and physical understanding of general audiences presented great difficulties. The author adopted a quasi-historical approach of presentation, illustrating each step with experiments and diagrams. In the printed version, he used a simple mathematical formalism, not exceeding the level in the syllabus of the lower grades of high school. In this sense, this textbook is different from the other books mentioned in this chapter because it takes the middle road between a popular account of the theory and a mathematical presentation of its ideas and implications.[29]

August Kopff. *The Mathematical Theory of Relativity*. This book is based on lectures at Heidelberg University in the winter semester of 1919–1920 and the summer semester of 1920. The author remarks in the introduction:

> That an astronomer has undertaken it to write such an introduction should not be a cause for astonishment. Because the theory of relativity is as much a concern of

> astronomy as it is of physics and not only because the demonstrations of its validity are currently undertaken just in the astronomical domain.[30]

Ludwik Silberstein. *The Theory of General Relativity and Gravitation*. This book is based on a series of fifteen lectures and two colloquia given by Silberstein in January 1921 at the University of Toronto. The author writes in the introduction:

> The book is felt to be far from being complete. But as it is, it is hoped that it will give the reader a good insight into the guiding spirit of Einstein's general relativity and gravitation theory and enable him to follow without serious difficulties the deeper investigations and the more special and extended developments given in the large and rapidly growing number of papers on the subject. Some of my readers will miss, perhaps, in this volume the enthusiastic tone which usually permeates the books and pamphlets that have been written on the subject (with a notable exception of Einstein's own writings). Yet the author is the last man to be blind to the admirable boldness and the severe architectonic beauty of Einstein's theory. But it has seemed that beauties of such a kind are rather enhanced than obscured by the adoption of a sober tone and an apparently cold form of presentation.[31]

H. ANTI-RELATIVITY LITERATURE

Numerous publications objecting to the theory of relativity, attempting to refute its validity, and criticizing, sometimes viciously, Einstein himself appeared in the 1920s. Most of them were published in the format of short booklets. They were written by physicists, academics of other disciplines, and professionals such as engineers, physicians, and intellectuals. The tensions between this kind of publication and the specialist research literature made their mark on the cultural and political arena and also provided occasions for Einstein to explain and defend his theory against critics of different categories. The anti-relativity movement in Germany and its various manifestations have been extensively analyzed in a book by Milena Wazeck, *Einstein's Opponents*.[32]

Many of the arguments and objections against the theory of relativity were formulated within the framework and language of physical discourse. Some accused the theory of relativity as being only a mathematical construct, without experimental validation; others argued that the ether as a medium necessary for carrying the electromagnetic field oscillation could not be discarded; still others, and actually most of them, claimed that relativity theory contradicted well-established concepts of space and time; some authors accepted the special theory of relativity, or parts of it, but rejected the general theory. Such and similar arguments constitute the main content of all the anti-relativity books and pamphlets.

In addition, all these objections were based on a shared understanding of the nature of science. Science should be comprehensible and simple, while the mathematical complexity of modern physics was perceived by Einstein's opponents as being alien to physical reality and as reflecting a decline of the traditional search for natural truth. This attitude was rooted in a history that had produced this kind of literature outside academic spheres long before Einstein. The increasing significance of science and technology in the second half of the nineteenth century, in fact, went hand in hand with the rise of the popularization of science. In this context, the role of science as a foundation for constructing world

views that had essentially been lost in the context of an ever-more-specialized professional science was revitalized and sometimes found its expression in quasi-religious ideologies. This situation also created a context for the development of world views based on scientific knowledge but critical of academic science by individuals who felt professionally competent to express such views, such as engineers or physicians. But many of them thought they had been excluded from modernization processes, embodied in the rapid development of ever more abstract scientific theories that seemed to be in conflict with commonsense. When Einstein's theories of relativity, which combined seemingly intuitive arguments about elementary measurement operations with complex mathematics, entered the stage, they were immediately seen as competitors, offering a natural target for criticism. For Einstein's opponents, the distortion of commonsense and the incomprehensibility of modern physics were particularly embodied in the general theory of relativity.[33]

A book written by the physician Karl Vogtherr was marketed with a blunt statement on its cover: "New! A generally comprehensible defense of common sense against Einstein's attacks!"[34] Einstein is personally depicted as the threat to commonsense and to the traditional conception of the goal and nature of scientific research. This is a common feature of the anti-relativity literature. Let us mention here a few of the many examples to give the reader an impression of this phenomenon, at least, from the titles:

- Gustav Richter-Bozen (1921) *Kritik der Relativitätstheorie Einstein's* (Leipzig: O. Hillmann) [Critique of Einstein's relativity theory]
- Rudolf Weinmann (1922) *Gegen Einsteins Relativierung von Zeit und Raum: gemeinverständlich* (Munich: R. Oldenbourg) [Against Einstein's relativization of time and space: a popular account]
- Rudolf Weinmann (1923) *Anti-Einstein* (Leipzig: O. Hillmann)
- Rudolf Weinmann (1925) *Widersprüche und Selbstwidersprüche der Relativitätstheorie* (Leipzig: O. Hillmann) [Contradictions and self-contradictions of relativity theory]
- Ernst Gehrcke (1924) *Die Massensuggestion der Relativitatstheorie* (Berlin H. Meusser) [The suggestion to the masses of relativity theory]

The publisher O. Hillmann in Leipzig specialized in anti-relativity literature. On the back cover of every book published by this company was an announcement that the company would send to anybody, upon request, a catalogue of its anti-Einstein literature.

A typical summary of the position of Einstein's opponents is given in the last paragraph of Rudolf Weinmann's 1925 booklet:

> Reason is altogether against the relativization of space and time; experiment and observation, by themselves never capable of shattering logic, do in no way enforce such an enterprise. Experience could because of natural-human constraints never lead to a confirmation. Economy of thinking requires the absoluteness of the space and time distance. Otherwise there is no end to contradictions and confusion, otherwise the—legitimate—relativity of all measuring rods is suspended, and nothing will be changed and gained other than words, which are without coverage by imaginable and conceivable content.

Weinmann was an initiator and coeditor of *100 Authors against Einstein*.[35] This book was published as a protest against the marginalization of the anti-relativity arguments.

Its main goal was to show that Einstein's opponents composed a large group that had to be taken seriously. The book actually contains not one hundred but twenty-eight original contributions and quotations from published literature. The introduction sets the tone:

> The purpose of this publication is to present an overview of the number and weight of the opponents and counterarguments against the terror of the Einsteinians. The purpose is to further the enlightenment of the public and the clarification of the problems in question.

A prominent player in the anti-Einstein movement was the experimental physicist Ernst Gehrcke, who aimed to reveal the theory of relativity as a "suggestion to the masses" disseminated by propaganda in the daily newspapers. His booklet *Die Massensuggestion der Relativitätstheorie* (1924) is based on an extensive newspaper article collection. To acquire the articles, Gehrcke subscribed to clipping services that sent him all articles containing the keywords "Einstein" or "relativity."[36] The opposition to Einstein and to his theory of relativity cannot be separated from the cultural, social, and political climate of the Weimar Republic. The physicist and historian of science Hubert Goenner argues that the anti-Einstein campaign accompanied the rise of nationalism and national political parties in Germany. He claims that the anti-Einstein campaign was organized, and the physicists and others involved in it were used primarily to gather support for one of the German right-wing parties, the Deutschnationale Volkspartei.[37]

The best-known, and in later years most vocal, Einstein's opponent was the Nobel laureate Philipp Lenard. His first publications on the theory of relativity are from 1918 and 1921.[38] In these publications, Lenard emphasized that he had no reservations against "the principle of relativity in the original form," namely, the special theory of relativity, but was critical of the general theory. These publications do not yet reveal any personal enmity against Einstein. However, their public debate in Bad Nauheim, in September 1920, was already marked by a sharp personal confrontation. Lenard insisted on comprehensibility, to which Einstein responded:

> I would like to say that what man considers comprehensible and what he does not has changed. The view of comprehensibility is to some extent a function of time. I think that physics is conceptual, not comprehensible.[39]

In later years, with the rise to power of the National Socialists, Lenard became increasingly infected with hate against Einstein, which was expressed in insulting anti-Semitic terms. He referred to Einstein as the "relativity Jew" and to modern physics as "Jew spirited." This attitude is demonstrated blatantly in Lenard's *Deutsche Physik*,[40] which distinguishes between German and Jewish physics, the latter being based on dogmatic spirit and damaging to German science.[41]

I. SUMMARY

In this chapter, we have listed and described most, but not all, of the many monographs and textbooks published during the period discussed in our book. The role of these early textbooks was more than to simply serve as didactic tools in achieving the curricula goals

of teaching institutions. They were read not only by students but also by their professors and by the key players in the field, who had to keep up with the latest developments. In some of these books, new ideas and proposals of the authors were fully presented for the first time. As part of the formative years, these books do not present a cleaned-up version of established scientific paradigms, as is typically the nature of common pedagogical texts. Instead, they provide an important source of inquiry for historians of science and inspiring reading material for anyone interested in the evolution of ideas.

The textbooks were indeed part of an exciting intellectual process from which relativity theory emerged and were a tool for both teaching and research. They connected the brand-new theory to the established canon of physical knowledge, an integration that eventually changed both the canon and the understanding of the theory itself. The textbooks thus served as a vehicle for an integration of knowledge that at the same time was a transformation of the structures of physical knowledge. They significantly influenced the future development of Einstein's theory, especially as the years considered here were followed by a period of marginalization of general relativity, as well as of political and human disasters that were foreshadowed in the anti-Einstein literature reviewed above. These disasters interrupted the steady accumulation of knowledge documented in the textbooks. They also represent a kind of message in a bottle for those who took up the exploration of the theory more vigorously during its renaissance in the 1950s.[42]

NOTES

1. Arnold Sommerfeld, preface to Wolfgang Pauli, *The Theory of Relativity*, trans. G. Field (New York, Pergamon Press; Bombay: B. I. Publications, 1958), p. xi.
2. Erwin Freundlich, *Die Grundlagen der Einsteinschen Gravitationstheorie: Mit e. Vorw. von Albert Einstein* (Berlin: Springer, 1916); trans. Henry L. Brose as *The Foundations of Einstein's Theory of Gravitation* (New York: G. E. Stechert, 1922), p. 42.
3. Freundlich, *Foundations of Einstein's Theory of Gravitation.*
4. Einstein to Erwin Freundlich, 19 September 1919, CPAE vol. 9, Doc. 106.
5. Einstein to Michele Besso, 3 January 1916, CPAE vol. 8, Doc. 178.
6. See Gutfreund and Renn, *Relativity: The Special and the General Theory*.
7. Ibid., p. 10.
8. Einstein to Michele Besso, 9 March 1917, CPAE vol. 8, Doc. 306.
9. Einstein, *On the Special and General Theory of Relativity*, CPAE vol. 6, Doc. 42.
10. See the preface to the first edition: Hermann Weyl, *Raum · Zeit · Materie: Vorlesungen über allgemeine Relativitätstheorie* (Heidelberg: Springer, 1918).
11. See the discussion in chapter 8, pp. 126–127.
12. Albert Einstein, "Review of Hermann Weyl, *Space–Time–Matter: Lectures on General Relativity*," CPAE vol. 7, Doc. 10.
13. See the discussion in chapter 8, pp. 126–127.
14. Eddington, *Space, Time and Gravitation*, p. v.
15. Eddington, *Mathematical Theory of Relativity*, p. v.
16. Ibid., p. 240.
17. Max Laue, *Das Relativitätsprinzip* (Braunschweig: Vieweg, 1911); Max von Laue, *Die Relativitätstheorie. Erster Band. Das Relativitätsprinzip der Lorentztransformation* (1921); and *Die Relativitätstheorie. Zweiter Band. Die Allgemeine Relativitätstheorie und Einsteins Lehre von der Schwerkraft* (1921).
18. On von Laue and relativistic continuum mechanics, see Michel Janssen and Matthew Mecklenburg, "From Classical to Relativistic Mechanics: Electromagnetic Models of the Electron," in *Interactions: Mathematics, Physics and Philosophy, 1860–1930*, ed. V. F. Hendricks, K. F. Jørgensen, J. Lützen, and S. A. Pedersen (Berlin: Springer, 2007), pp. 65–134. For a discussion of the influence of von Laue's book on Einstein's thinking, see Rowe, "Max von Laue's Role in the Relativity Revolution."

19. Max von Laue to Moritz Schlick, Eggishorn, 19 August 1913, Inv.-Nr. 108/Lau-15, Noord-Hollands Archief Haarlem (NL).
20. Max von Laue, *Die Relativitaetstheorie*, Zweiter Band, p. viii. Authors' translation.
21. Ernst Cassirer, *Substance and Function and Einstein's Theory of Relativity* (Chicago: Open Court, 1923).
22. Ibid., p. vi. Authors' translation.
23. Arnold Sommerfeld, preface to Pauli, *Theory of Relativity*, p. xi (Pergamon Press ed.).
24. Albert Einstein, "Review of Wolfgang Pauli, *The Theory of Relativity*," CPAE vol. 13, Doc. 62.
25. Pauli, "Relativitätstheorie"; *Theory of Relativity*, p. 206.
26. Hans Thirring, *Die Idee der Relativitaetstheorie* (Berlin: Springer, 1921); trans. Rhoda A. B. Russell as *The Ideas of Einstein's Theory: The Theory of Relativity in Simple Language* (London: Methuen, 1921)
27. Hans Thirring, *Ideas of Einstein's Theory*, pp. vi–vii.
28. Hans Thirring, *L'Idée de la théorie de la relativité* (Paris: Gauthier Villars, 1923). Excerpt in CPAE vol. 13, Doc. 234.
29. Max Born, *Die Relativitätstheorie Einsteins und ihre physikalischen Grundlagen* (Berlin: Springer, 1920); trans. Henry L. Brose as *Einstein's Theory of Relativity* (New York: Dutton, 1922).
30. August Kopff, *Grundzuege der Einsteinschen Relativitaetstheorie* (Leipzig: Hirzel, 1921); trans. H. Levy as *The Mathematical Theory of Relativity* (London: Methuen, 1923). For more on Kopff and Einstein, see Roland Wielen and Ute Wielen, *August Kopff, die Relativitätstheorie, und zwei Briefe Albert Einsteins an Kopff im Archiv des Astronomischen Rechen-Instituts* (Heidelberg: Astronomisches Rechen-Institut Zentrum für Astronomie Universität Heidelberg, 2003), http://archiv.ub.uni-heidelberg.de/volltextserver/15653/1/Kopff_Einstein_1.pdf.
31. Ludwik Silberstein, *The Theory of General Relativity and Gravitation* (New York: Van Nostrand, 1922).
32. Wazeck, *Einstein's Opponents*.
33. Ibid., pp. 147–153.
34. Karl Vogtherr, *Wohin führt die Relativitätstheorie? Kritische Betrachtungen vom physikalischen und erkenntnistheoretischen Standpunkt aus* [Where does the theory of relativity lead? Critical reflections from the standpoint of physics and epistemology] (Leipzig: O. Hillmann, 1923).
35. Hans Israel, Erich Ruckhaber, and Rudolf Weinmann, eds., *Hundert Autoren gegen Einstein* (Leipzig: R. Voigtländer, 1931). For historical discussion, see Goenner, "The Reaction to Relativity Theory."
36. This collection is held by the Max Planck Institute for the History of Science in Berlin : http://echo.mpiwg-berlin.mpg.de/content/modernphysics/gehrcke.
37. Hubert Goenner, "The Reaction to Relativity Theory I: The Anti-Einstein Campaign in Germany in 1920," *Science in Context* 6 (1993): 107–133.
38. Philip Lenard, *Über Relativitätsprinzip, Äther, Gravitation* (Leipzig: Hirzel, 1918); and *Über Äther und Uräther* (Leipzig: Hirzel, 1921).
39. "Discussions of Lectures in Bad Neuheim," CPAE vol. 7 (German version), Doc. 46, p. 354.
40. Philipp Lenard, *Deutsche Physik in vier Bänden* (Munich: Lehmanns, 1936).
41. Andreas Kleinert, "Philipp Lenard and Johannes Stark: Two Nobel Laureates against Einstein," in *Albert Einstein: One Hundred Authors for Einstein*, ed. J. Renn (Weinheim: Wiley-VCH, 2005), p. 228.
42. See the discussion in chapter 10. For a comparison with the role of textbooks in the emergence of quantum theory, see Massimiliano Badino and Jaume Navarro (eds.), *Research and Pedagogy: A History of Quantum Physics through Its Textbooks* (Berlin: Edition Open Access, 2013), http://edition-open-access.de/studies/2/index.html.

10
BEYOND THE FORMATIVE YEARS

The commentaries in this book cover the early years of general relativity, when its epistemic status was uncertain in many respects: from the understanding of its physical implications via the understanding of its mathematical apparatus, to its philosophical interpretation, which constituted a major issue for many contemporary scientists, as it was for Einstein himself. The correspondence about gravitational waves and approximate solutions between Einstein, Karl Schwarzschild, and Willem de Sitter is one example of how unclear and unexplored the connections between the theory and its physical consequences were in those early days.

And if Einstein's quest for the interpretation of his theory was indeed philosophy, it was a philosophy in flux. But it was actually part of a more comprehensive effort of conceptual reflection that cannot really be confined to a clearly delineated domain. These are the true reasons behind the difficulty in pinpointing Einstein's position, its evolving character, and its comprehensive, albeit somewhat eclectic, scope, which was not limited to what by today's standards would pass as physics or philosophy sensu strictu.

Evidently, since these formative years of general relativity, at a time when the boundaries between physics, philosophy, psychology, and history were more fluent and less well defined than they are today, we have witnessed what amounts to a veritable split of scientific rationality along disciplinary fault lines, a split that makes it even more worthwhile to take a closer look at Einstein's struggles with the meaning of his own theory. His achievement is indeed best characterized by the very title of the English version of the book in which he published his Princeton lectures: his quest was to understand the "meaning of relativity," in the sense of anchoring his novel scientific insights in the legacy of human thinking, thus imbuing them with meaning beyond that of a mere formalism that would allow some minute effects to be calculated more accurately than Newton's theory could do.

After Eddington's observations of the bending of light, there was a growing interest in general relativity, bringing more and more scientists into the field. This interest lasted throughout the 1920s. This is the period to which we refer as the formative years. Following this period, the theory underwent a process of marginalization within the field of physics. Jean Eisenstaedt characterized these years as the "low-water-mark" period, lasting until the 1950s.[1] During this period most physicists considered general relativity as a highly formalistic theory, providing only small corrections to the Newtonian picture.

Only a limited number of theoretical physicists worked on the theory while the majority focused on the development of quantum theory, which had much stronger connections with experimental activities and possible technological applications. And even those research projects concerned with general relativity during the low-water-mark period often aimed at moving beyond general relativity and in any case had few connections to empirical investigations.

The only empirical domain in which general relativity had some impact beyond small corrections to the Newtonian picture was cosmology. As Eisenstaedt observed, "cosmology was an intellectual laboratory" that served to explore the consequences of general relativity well beyond its role for deriving corrections to Newton's theory of gravitation.[2] He has turned our attention to the role of the American chemical physicist Richard C. Tolman as an important conduit between the formative years and the renaissance of general relativity, including the early work on stellar collapse.[3] In chapter 5 we discussed Tolman's encounter with Einstein and the friendship they struck up. Tolman had not only written the first introduction to special relativity in the United States, in 1934 he also published a unique textbook entitled *Relativity, Thermodynamics and Cosmology*, in which he gave a comprehensive overview of the field with an emphasis on the physical aspects of the theory and also presented his own investigations into the extension of thermodynamics to general relativity, as well as a detailed treatment of non-static models of the universe and their relation to cosmology and observational knowledge. Einstein wrote an enthusiastic review of Tolman's book that captures the role it was to play in carrying the message of relativity to future generations:

> The author thus succeeded, in my opinion, to offer a systematic treatment of all those methods and results of the theory of relativity which seem to be qualified to enter into any later theory that penetrates more deeply into the mechanism of what is happening.

While taking the occasion to once more abdicate the cosmological constant as an arbitrary and unnecessary addition to his theory, Einstein concluded by stressing the significance of Tolman's courageous exploration of cosmology:

> What I find particularly commendable about Tolman's book is the exhaustive treatment of the laws to be expected for the nebulae according to the theory; because these seem to be primarily suited to complete our knowledge about the structure of the space-time continuum.

But even in cosmology, only certain aspects of general relativity were considered useful in specific contexts, while they were not investigated from the perspective of an encompassing research program with general relativity at its center.

These dispersed research traditions, however, kept the interest in the theory alive until the early 1960s, when the theory returned to the mainstream of physics. Clifford Will labeled this period as the "renaissance of general relativity."[4] But as we have emphasized, its beginnings can actually be traced to the mid-1950s. Around 1955, a number of research centers focusing on aspects of general relativity were active in different parts of

the world. One crucial element in the dynamics of the renaissance of general relativity was the establishment of a stable tradition of a long postdoctoral education. Another essential factor was the explicit attempts at building a community integrating and transforming the existing diverse and only loosely connected research agendas.[5]

As we have discussed in chapter 6 (p. 103), one important starting point for this community-building was a conference held 1955 in Bern that had been organized with the intention to celebrate the fiftieth anniversary of special relativity but concerned mostly topics and problems of general relativity. The Bern conference provided an occasion to recognize links between the existing dispersed research traditions and to draw consequences from this recognition for future research and collaborations. This realization led to the organization of further important meetings, such as the Chapel Hill conference in 1957 and the Royaumont conference in 1959, which then became the starting point for a long-lasting tradition of conferences, as well as for the creation of international bodies and collaborations.[6] These activities, together with the astrophysical discoveries of the 1960s, such as those of quasars and of the cosmic microwave background, prepared the ground for what Kip Thorne has called the "golden age" of general relativity in the 1960s and 1970s.[7]

Many of the achievements from the formative years had meanwhile been forgotten and had to be rediscovered. As the historian and relativist John Stachel once observed, there is a remarkable number of results in general relativity that were found two or even more times, testifying to its fragility as a field and as a tradition.[8] Still, neither the renaissance nor the golden years would have been possible without the broad basis for general relativity that was built during the formative years. This basis encompasses the early work of astronomers and mathematicians, the philosophical struggles surrounding its interpretation, the first efforts to integrate Einstein's ideas in the larger field of physics, and the first textbooks elaborating the theory and making it accessible to the next generation. All of these achievements made the later flourishing of general relativity possible. It is to this forgotten legacy of the formative years that we dedicate this book.

NOTES

1. See Eisenstaedt, "La relativité générale à l'herétiage: 1925–1955"; "Low Water Mark of General Relativity"; *The Curious History of Relativity*"; see also chapter 6.
2. Jean Eisenstaedt, "Cosmology: A Space for Thought on General Relativity," in *Foundations of Big Bang Cosmology*, ed. F. W. Meyerstein (Singapore: World Scientific, 1989), pp. 271–295, p. 264. See also Eisenstaedt, "Lemaître and the Schwarzschild solution."
3. Richard C. Tolman, *Relativity, Thermodynamics and Cosmology* (Oxford: Clarendon Press, 1934).
4. Clifford Will, *Was Einstein Right?*, esp. pp. 3–18.
5. See Blum et al., "Reinvention of General Relativity"; "The Renaissance of General Relativity"; "The Renaissance of Einstein's Theory of Gravitation"; Roberto Lalli, *Building the General Relativity and Gravitation Community During the Cold War* (Cham: Springer, 2017).
6. DeWitt and Rickles, *Role of Gravitation in Physics*; Jean Charles Marcel Jacquot, *Les Théâtres d'Asie. Études de Jeannine Auboyer [and others] . . . réunies et présentées par Jean Jacquot.* [Proceedings of the Conférences du Théâtre des nations, 1958–1959, and of the Journées d'herétudes de Royaumont, 28 May–1 June 1959.] (Paris: Government Publication, 1961).
7. Thorne, *Black Holes and Time Warps*.
8. John Stachel, "The Cauchy Problem in General Relativity—The Early Years," in *Studies in the History of General Relativity*, p. 407.

PART III

EINSTEIN'S BOOK WITH THE APPENDIXES

SPACE AND TIME IN PRE-RELATIVITY PHYSICS

THE theory of relativity is intimately connected with the theory of space and time. I shall therefore begin with a brief investigation of the origin of our ideas of space and time, although in doing so I know that I introduce a controversial subject. The object of all science, whether natural science or psychology, is to co-ordinate our experiences and to bring them into a logical system. How are our customary ideas of space and time related to the character of our experiences?

The experiences of an individual appear to us arranged in a series of events; in this series the single events which we remember appear to be ordered according to the criterion of "earlier" and "later," which cannot be analysed further. There exists, therefore, for the individual, an I-time, or subjective time. This in itself is not measurable. I can, indeed, associate numbers with the events, in such a way that a greater number is associated with the later event than with an earlier one; but the nature of this association may be quite arbitrary. This association I can define by means of a clock by comparing the order of events furnished by the clock with the order of the given series of events. We understand by a clock something which provides a series of events which can be counted, and which has other properties of which we shall speak later.

By the aid of language different individuals can, to a certain extent, compare their experiences. Then it turns out

that certain sense perceptions of different individuals correspond to each other, while for other sense perceptions no such correspondence can be established. We are accustomed to regard as real those sense perceptions which are common to different individuals, and which therefore are, in a measure, impersonal. The natural sciences, and in particular, the most fundamental of them, physics, deal with such sense perceptions. The conception of physical bodies, in particular of rigid bodies, is a relatively constant complex of such sense perceptions. A clock is also a body, or a system, in the same sense, with the additional property that the series of events which it counts is formed of elements all of which can be regarded as equal.

The only justification for our concepts and system of concepts is that they serve to represent the complex of our experiences; beyond this they have no legitimacy. I am convinced that the philosophers have had a harmful effect upon the progress of scientific thinking in removing certain fundamental concepts from the domain of empiricism, where they are under our control, to the intangible heights of the *a priori*. For even if it should appear that the universe of ideas cannot be deduced from experience by logical means, but is, in a sense, a creation of the human mind, without which no science is possible, nevertheless this universe of ideas is just as little independent of the nature of our experiences as clothes are of the form of the human body. This is particularly true of our concepts of time and space, which physicists have been obliged by the facts to bring down from the Olympus of the *a priori* in order to adjust them and put them in a serviceable condition.

We now come to our concepts and judgments of space. It is essential here also to pay strict attention to the rela-

PRE-RELATIVITY PHYSICS

tion of experience to our concepts. It seems to me that Poincaré clearly recognized the truth in the account he gave in his book, "La Science et l'Hypothese." Among all the changes which we can perceive in a rigid body those which can be cancelled by a voluntary motion of our body are marked by their simplicity; Poincaré calls these, changes in position. By means of simple changes in position we can bring two bodies into contact. The theorems of congruence, fundamental in geometry, have to do with the laws that govern such changes in position. For the concept of space the following seems essential. We can form new bodies by bringing bodies B, C, . . . up to body A; we say that we *continue* body A. We can continue body A in such a way that it comes into contact with any other body, X. The ensemble of all continuations of body A we can designate as the "space of the body A." Then it is true that all bodies are in the "space of the (arbitrarily chosen) body A." In this sense we cannot speak of space in the abstract, but only of the "space belonging to a body A." The earth's crust plays such a dominant rôle in our daily life in judging the relative positions of bodies that it has led to an abstract conception of space which certainly cannot be defended. In order to free ourselves from this fatal error we shall speak only of "bodies of reference," or "space of reference." It was only through the theory of general relativity that refinement of these concepts became necessary, as we shall see later.

I shall not go into detail concerning those properties of the space of reference which lead to our conceiving points as elements of space, and space as a continuum. Nor shall I attempt to analyse further the properties of space which justify the conception of continuous series of points, or lines. If these concepts are assumed, together

with their relation to the solid bodies of experience, then it is easy to say what we mean by the three-dimensionality of space; to each point three numbers, x_1, x_2, x_3 (co-ordinates), may be associated, in such a way that this association is uniquely reciprocal, and that x_1, x_2, and x_3 vary continuously when the point describes a continuous series of points (a line).

It is assumed in pre-relativity physics that the laws of the configuration of ideal rigid bodies are consistent with Euclidean geometry. What this means may be expressed as follows: Two points marked on a rigid body form an *interval*. Such an interval can be oriented at rest, relatively to our space of reference, in a multiplicity of ways. If, now, the points of this space can be referred to co-ordinates x_1, x_2, x_3, in such a way that the differences of the co-ordinates, Δx_1, Δx_2, Δx_3, of the two ends of the interval furnish the same sum of squares,

$$s^2 = \Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 \tag{1}$$

for every orientation of the interval, then the space of reference is called Euclidean, and the co-ordinates Cartesian.* It is sufficient, indeed, to make this assumption in the limit for an infinitely small interval. Involved in this assumption there are some which are rather less special, to which we must call attention on account of their fundamental significance. In the first place, it is assumed that one can move an ideal rigid body in an arbitrary manner. In the second place, it is assumed that the behaviour of ideal rigid bodies towards orientation is independent of the material of the bodies and their changes of position, in the sense that if two intervals can

* This relation must hold for an arbitrary choice of the origin and of the direction (ratios $\Delta x_1 : \Delta x_2 : \Delta x_3$) of the interval.

once be brought into coincidence, they can always and everywhere be brought into coincidence. Both of these assumptions, which are of fundamental importance for geometry and especially for physical measurements, naturally arise from experience; in the theory of general relativity their validity needs to be assumed only for bodies and spaces of reference which are infinitely small compared to astronomical dimensions.

The quantity s we call the length of the interval. In order that this may be uniquely determined it is necessary to fix arbitrarily the length of a definite interval; for example, we can put it equal to 1 (unit of length). Then the lengths of all other intervals may be determined. If we make the x_ν linearly dependent upon a parameter λ,

$$x_\nu = a_\nu + \lambda b_\nu,$$

we obtain a line which has all the properties of the straight lines of the Euclidean geometry. In particular, it easily follows that by laying off n times the interval s upon a straight line, an interval of length $n \cdot s$ is obtained. A length, therefore, means the result of a measurement carried out along a straight line by means of a unit measuring rod. It has a significance which is as independent of the system of co-ordinates as that of a straight line, as will appear in the sequel.

We come now to a train of thought which plays an analogous rôle in the theories of special and general relativity. We ask the question: besides the Cartesian co-ordinates which we have used are there other equivalent co-ordinates? An interval has a physical meaning which is independent of the choice of co-ordinates; and so has the spherical surface which we obtain as the locus of the end points of all equal intervals that we lay off from an

arbitrary point of our space of reference. If x_ν as well as x'_ν (ν from 1 to 3) are Cartesian co-ordinates of our space of reference, then the spherical surface will be expressed in our two systems of co-ordinates by the equations

$$\sum \Delta x_\nu^2 = \text{const.} \tag{2}$$

$$\sum \Delta x'^2_\nu = \text{const.} \tag{2a}$$

How must the x'_ν be expressed in terms of the x_ν in order that equations (2) and (2a) may be equivalent to each other? Regarding the x'_ν expressed as functions of the x_ν, we can write, by Taylor's theorem, for small values of the Δx_ν,

$$\Delta x'_\nu = \sum_\alpha \frac{\partial x'_\nu}{\partial x_\alpha} \Delta x_\alpha + \frac{1}{2} \sum_{\alpha\beta} \frac{\partial^2 x'_\nu}{\partial x_\alpha \partial x_\beta} \Delta x_\alpha \Delta x_\beta \ . \ . \ .$$

If we substitute (2a) in this equation and compare with (1), we see that the x'_ν must be linear functions of the x_ν. If we therefore put

$$x'_\nu = \alpha_\nu + \sum_\alpha b_{\nu\alpha} x_\alpha \tag{3}$$

or

$$\Delta x'_\nu = \sum_\alpha b_{\nu\alpha} \Delta x_\alpha \tag{3a}$$

then the equivalence of equations (2) and (2a) is expressed in the form

$$\sum \Delta x'^2_\nu = \lambda \sum \Delta x_\nu^2 \ (\lambda \text{ independent of } \Delta x_\nu) \tag{2b}$$

It therefore follows that λ must be a constant. If we put $\lambda = 1$, (2b) and (3a) furnish the conditions

$$\sum_\nu b_{\nu\alpha} b_{\nu\beta} = \delta_{\alpha\beta} \tag{4}$$

in which $\delta_{\alpha\beta} = 1$, or $\delta_{\alpha\beta} = 0$, according as $\alpha = \beta$ or $\alpha \neq \beta$. The conditions (4) are called the conditions of orthogonality, and the transformations (3), (4), linear orthogonal transformations. If we stipulate that $s^2 = \sum \Delta x_\nu^2$ shall be equal to the square of the length in every system of co-ordinates, and if we always measure with the same unit scale, then λ must be equal to 1. Therefore the linear orthogonal transformations are the only ones by means of which we can pass from one Cartesian system of co-ordinates in our space of reference to another. We see that in applying such transformations the equations of a straight line become equations of a straight line. Reversing equations (3a) by multiplying both sides by $b_{\nu\beta}$ and summing for all the ν's, we obtain

$$(5) \qquad \sum b_{\nu\beta}\Delta x'_\nu = \sum_{\nu\alpha} b_{\nu\alpha}b_{\nu\beta}\Delta x_\alpha = \sum_{\alpha} \delta_{\alpha\beta}\Delta x_\alpha = \Delta x_\beta$$

The same coefficients, b, also determine the inverse substitution of Δx_ν. Geometrically, $b_{\nu\alpha}$ is the cosine of the angle between the x'_ν axis and the x_α axis.

To sum up, we can say that in the Euclidean geometry there are (in a given space of reference) preferred systems of co-ordinates, the Cartesian systems, which transform into each other by linear orthogonal transformations. The distance s between two points of our space of reference, measured by a measuring rod, is expressed in such co-ordinates in a particularly simple manner. The whole of geometry may be founded upon this conception of distance. In the present treatment, geometry is related to actual things (rigid bodies), and its theorems are statements concerning the behaviour of these things, which may prove to be true or false.

PRE-RELATIVITY PHYSICS

One is ordinarily accustomed to study geometry divorced from any relation between its concepts and experience. There are advantages in isolating that which is purely logical and independent of what is, in principle, incomplete empiricism. This is satisfactory to the pure mathematician. He is satisfied if he can deduce his theorems from axioms correctly, that is, without errors of logic. The questions as to whether Euclidean geometry is true or not does not concern him. But for our purpose it is necessary to associate the fundamental concepts of geometry with natural objects; without such an association geometry is worthless for the physicist. The physicist is concerned with the question as to whether the theorems of geometry are true or not. That Euclidean geometry, from this point of view, affirms something more than the mere deductions derived logically from definitions may be seen from the following simple consideration.

Between n points of space there are $\frac{n(n-1)}{2}$ distances, $s_{\mu\nu}$; between these and the $3n$ co-ordinates we have the relations

$$s_{\mu\nu}^2 = (x_{1(\mu)} - x_{1(\nu)})^2 + (x_{2(\mu)} - x_{2(\nu)})^2 + \ldots$$

From these $\frac{n(n-1)}{2}$ equations the $3n$ co-ordinates may be eliminated, and from this elimination at least $\frac{n(n-1)}{2} - 3n$ equations in the $s_{\mu\nu}$ will result.* Since the $s_{\mu\nu}$ are measurable quantities, and by definition are independent of each other, these relations between the $s_{\mu\nu}$ are not necessary *a priori*.

* In reality there are $\frac{n(n-1)}{2} - 3n + 6$ equations.

PRE-RELATIVITY PHYSICS

From the foregoing it is evident that the equations of transformation (3), (4) have a fundamental significance in Euclidean geometry, in that they govern the transformation from one Cartesian system of co-ordinates to another. The Cartesian systems of co-ordinates are characterized by the property that in them the measurable distance between two points, s, is expressed by the equation

$$s^2 = \sum \Delta x_\nu^2.$$

If $K_{(x_\nu)}$ and $K'_{(x_\nu)}$ are two Cartesian systems of co-ordinates, then

$$\sum \Delta x_\nu^2 = \sum \Delta x'_\nu{}^2.$$

The right-hand side is identically equal to the left-hand side on account of the equations of the linear orthogonal transformation, and the right-hand side differs from the left-hand side only in that the x_ν are replaced by the x'_ν. This is expressed by the statement that $\sum \Delta x_\nu^2$ is an invariant with respect to linear orthogonal transformations. It is evident that in the Euclidean geometry only such, and all such, quantities have an objective significance, independent of the particular choice of the Cartesian co-ordinates, as can be expressed by an invariant with respect to linear orthogonal transformations. This is the reason that the theory of invariants, which has to do with the laws that govern the form of invariants, is so important for analytical geometry.

As a second example of a geometrical invariant, consider a volume. This is expressed by

$$V = \int\int\int dx_1 dx_2 dx_3.$$

PRE-RELATIVITY PHYSICS

By means of Jacobi's theorem we may write

$$\int\int\int dx'_1 dx'_2 dx'_3 = \int\int\int \frac{\partial(x'_1, x'_2, x'_3)}{\partial(x_1, x_2, x_3)}\, dx_1 dx_2 dx_3$$

where the integrand in the last integral is the functional determinant of the x'_ν with respect to the x_ν, and this by (3) is equal to the determinant $|b_{\mu\nu}|$ of the coefficients of substitution, $b_{\nu\alpha}$. If we form the determinant of the $\delta_{\mu\alpha}$ from equation (4), we obtain, by means of the theorem of multiplication of determinants,

$$(6) \qquad 1 = |\delta_{\alpha\beta}| = \left|\sum_\nu b_{\nu\alpha} b_{\nu\beta}\right| = |b_{\mu\nu}|^2;\ |b_{\mu\nu}| = \pm 1$$

If we limit ourselves to those transformations which have the determinant $+1$* (and only these arise from continuous variations of the systems of co-ordinates) then V is an invariant.

Invariants, however, are not the only forms by means of which we can give expression to the independence of the particular choice of the Cartesian co-ordinates. Vectors and tensors are other forms of expression. Let us express the fact that the point with the current co-ordinates x_ν lies upon a straight line. We have

$$x_\nu - A_\nu = \lambda B_\nu \ (\nu \text{ from 1 to 3}).$$

Without limiting the generality we can put

$$\sum B_\nu^2 = 1.$$

* There are thus two kinds of Cartesian systems which are designated as "right-handed" and "left-handed" systems. The difference between these is familiar to every physicist and engineer. It is interesting to note that these two kinds of systems cannot be defined geometrically, but only the contrast between them.

PRE-RELATIVITY PHYSICS

If we multiply the equations by $b_{\beta\nu}$ (compare (3a) and (5)) and sum for all the ν's, we get

$$x'_\beta - A'_\beta = \lambda B'_\beta$$

where we have written

$$B'_\beta = \sum_\nu b_{\beta\nu} B_\nu; \; A'_\beta = \sum_\nu b_{\beta\nu} A_\nu.$$

These are the equations of straight lines with respect to a second Cartesian system of co-ordinates K'. They have the same form as the equations with respect to the original system of co-ordinates. It is therefore evident that straight lines have a significance which is independent of the system of co-ordinates. Formally, this depends upon the fact that the quantities $(x_\nu - A_\nu) - \lambda B_\nu$ are transformed as the components of an interval, Δx_ν. The ensemble of three quantities, defined for every system of Cartesian co-ordinates, and which transform as the components of an interval, is called a vector. If the three components of a vector vanish for one system of Cartesian co-ordinates, they vanish for all systems, because the equations of transformation are homogeneous. We can thus get the meaning of the concept of a vector without referring to a geometrical representation. This behaviour of the equations of a straight line can be expressed by saying that the equation of a straight line is co-variant with respect to linear orthogonal transformations.

We shall now show briefly that there are geometrical entities which lead to the concept of tensors. Let P_0 be the centre of a surface of the second degree, P any point on the surface, and ξ_ν the projections of the interval P_0P upon the co-ordinate axes. Then the equation of the

surface is

$$\sum a_{\mu\nu}\xi_\mu\xi_\nu = 1.$$

In this, and in analogous cases, we shall omit the sign of summation, and understand that the summation is to be carried out for those indices that appear twice. We thus write the equation of the surface

$$a_{\mu\nu}\xi_\mu\xi_\nu = 1.$$

The quantities $a_{\mu\nu}$ determine the surface completely, for a given position of the centre, with respect to the chosen system of Cartesian co-ordinates. From the known law of transformation for the ξ_ν (3a) for linear orthogonal transformations, we easily find the law of transformation for the $a_{\mu\nu}$*:

$$a'_{\sigma\tau} = b_{\sigma\mu}b_{\tau\nu}a_{\mu\nu}.$$

This transformation is homogeneous and of the first degree in the $a_{\mu\nu}$. On account of this transformation, the $a_{\mu\nu}$ are called components of a tensor of the second rank (the latter on account of the double index). If all the components, $a_{\mu\nu}$, of a tensor with respect to any system of Cartesian co-ordinates vanish, they vanish with respect to every other Cartesian system. The form and the position of the surface of the second degree is described by this tensor (a).

Tensors of higher rank (number of indices) may be defined analytically. It is possible and advantageous to regard vectors as tensors of rank 1, and invariants (scalars) as tensors of rank 0. In this respect, the problem of the theory of invariants may be so formulated: according to

* The equation $a'_{\sigma\tau}\xi'_\sigma\xi'_\tau = 1$ may, by (5), be replaced by $a'_{\sigma\tau}b_{\mu\sigma}b_{\nu\tau}\xi_\sigma\xi_\tau = 1$, from which the result stated immediately follows.

what laws may new tensors be formed from given tensors? We shall consider these laws now, in order to be able to apply them later. We shall deal first only with the properties of tensors with respect to the transformation from one Cartesian system to another in the same space of reference, by means of linear orthogonal transformations. As the laws are wholly independent of the number of dimensions, we shall leave this number, n, indefinite at first.

Definition. If an object is defined with respect to every system of Cartesian co-ordinates in a space of reference of n dimensions by the n^α numbers $A_{\mu\nu\rho} \ldots$ (α = number of indices), then these numbers are the components of a tensor of rank α if the transformation law is

$$A'_{\mu'\nu'\rho'} \ldots = b_{\mu'\mu} b_{\nu'\nu} b_{\rho'\rho} \ldots A_{\mu\nu\rho} \ldots \tag{7}$$

Remark. From this definition it follows that

$$A_{\mu\nu\rho} \ldots B_\mu C_\nu D_\rho \ldots \tag{8}$$

is an invariant, provided that (B), (C), (D) . . . are vectors. Conversely, the tensor character of (A) may be inferred, if it is known that the expression (8) leads to an invariant for an arbitrary choice of the vectors (B), (C), etc.

Addition and Subtraction. By addition and subtraction of the corresponding components of tensors of the same rank, a tensor of equal rank results:

$$A_{\mu\nu\rho} \ldots \pm B_{\mu\nu\rho} \ldots = C_{\mu\nu\rho} \ldots \tag{9}$$

The proof follows from the definition of a tensor given above.

Multiplication. From a tensor of rank α and a tensor of rank β we may obtain a tensor of rank $\alpha + \beta$ by

multiplying all the components of the first tensor by all the components of the second tensor:

$$(10) \qquad T_{\mu\nu\rho \ldots \alpha\beta\gamma \ldots} = A_{\mu\nu\rho \ldots} B_{\alpha\beta\gamma \ldots}$$

Contraction. A tensor of rank $\alpha - 2$ may be obtained from one of rank α by putting two definite indices equal to each other and then summing for this single index:

$$(11) \qquad T_{\rho \ldots} = A_{\mu\mu\rho \ldots} \left(= \sum_{\mu} A_{\mu\mu\rho \ldots} \right)$$

The proof is

$$A'_{\mu\mu\rho \ldots} = b_{\mu\alpha} b_{\mu\beta} b_{\rho\gamma} \ldots A_{\alpha\beta\gamma \ldots} = \delta_{\alpha\beta} b_{\rho\gamma} \ldots A_{\alpha\beta\gamma \ldots} = b_{\rho\gamma} \ldots A_{\alpha\alpha\gamma \ldots}$$

In addition to these elementary rules of operation there is also the formation of tensors by differentiation ("Erweiterung"):

$$(12) \qquad T_{\mu\nu\rho \ldots \alpha} = \frac{\partial A_{\mu\nu\rho \ldots}}{\partial x_\alpha}$$

New tensors, in respect to linear orthogonal transformations, may be formed from tensors according to these rules of operation.

Symmetry Properties of Tensors. Tensors are called symmetrical or skew-symmetrical in respect to two of their indices, μ and ν, if both the components which result from interchanging the indices μ and ν are equal to each other or equal with opposite signs.

Condition for symmetry: $A_{\mu\nu\rho} = A_{\nu\mu\rho}$.

Condition for skew-symmetry: $A_{\mu\nu\rho} = -A_{\nu\mu\rho}$.

Theorem. The character of symmetry or skew-symmetry exists independently of the choice of co-ordinates, and in this lies its importance. The proof follows from the equation defining tensors.

PRE-RELATIVITY PHYSICS

Special Tensors.

I. The quantities $\delta_{\rho\sigma}$ (4) are tensor components (fundamental tensor).

Proof. If in the right-hand side of the equation of transformation $A'_{\mu\nu} = b_{\mu\alpha}b_{\nu\beta}A_{\alpha\beta}$, we substitute for $A_{\alpha\beta}$ the quantities $\delta_{\alpha\beta}$ (which are equal to 1 or 0 according as $\alpha = \beta$ or $\alpha \neq \beta$), we get

$$A'_{\mu\nu} = b_{\mu\alpha}b_{\nu\alpha} = \delta_{\mu\nu}.$$

The justification for the last sign of equality becomes evident if one applies (4) to the inverse substitution (5).

II. There is a tensor $(\delta_{\mu\nu\rho} \ldots)$ skew-symmetrical with respect to all pairs of indices, whose rank is equal to the number of dimensions, n, and whose components are equal to $+1$ or -1 according as $\mu\nu\rho \ldots$ is an even or odd permutation of $123 \ldots$

The proof follows with the aid of the theorem proved above $|b_{\rho\sigma}| = 1$.

These few simple theorems form the apparatus from the theory of invariants for building the equations of pre-relativity physics and the theory of special relativity.

We have seen that in pre-relativity physics, in order to specify relations in space, a body of reference, or a space of reference, is required, and, in addition, a Cartesian system of co-ordinates. We can fuse both these concepts into a single one by thinking of a Cartesian system of co-ordinates as a cubical frame-work formed of rods each of unit length. The co-ordinates of the lattice points of this frame are integral numbers. It follows from the fundamental relation

$$s^2 = \Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 \tag{13}$$

that the members of such a space-lattice are all of unit

length. To specify relations in time, we require in addition a standard clock placed, say, at the origin of our Cartesian system of co-ordinates or frame of reference. If an event takes place anywhere we can assign to it three co-ordinates, x_ν, and a time t, as soon as we have specified the time of the clock at the origin which is simultaneous with the event. We therefore give (hypothetically) an objective significance to the statement of the simultaneity of distant events, while previously we have been concerned only with the simultaneity of two experiences of an individual. The time so specified is at all events independent of the position of the system of co-ordinates in our space of reference, and is therefore an invariant with respect to the transformation (3).

It is postulated that the system of equations expressing the laws of pre-relativity physics is co-variant with respect to the transformation (3), as are the relations of Euclidean geometry. The isotropy and homogeneity of space is expressed in this way.* We shall now consider some of the more important equations of physics from this point of view.

The equations of motion of a material particle are

$$m\frac{d^2x_\nu}{dt^2} = X_\nu \tag{14}$$

(dx_ν) is a vector; dt, and therefore also $\frac{1}{dt}$, an invariant;

* The laws of physics could be expressed, even in case there were a preferred direction in space, in such a way as to be co-variant with respect to the transformation (3); but such an expression would in this case be unsuitable. If there were a preferred direction in space it would simplify the description of natural phenomena to orient the system of co-ordinates in a definite way with respect to this direction. But if, on the other hand, there is no unique direction in space it is not logical to formulate the laws of nature in such a way as to conceal the equivalence of systems of co-ordinates that are oriented differently. We shall meet with this point of view again in the theories of special and general relativity.

PRE-RELATIVITY PHYSICS

thus $\left(\frac{dx_\nu}{dt}\right)$ is a vector; in the same way it may be shown that $\left(\frac{d^2x_\nu}{dt^2}\right)$ is a vector. In general, the operation of differentiation with respect to time does not alter the tensor character. Since m is an invariant (tensor of rank 0), $\left(m\frac{d^2x_\nu}{dt^2}\right)$ is a vector, or tensor of rank 1 (by the theorem of the multiplication of tensors). If the force (X_ν) has a vector character, the same holds for the difference $\left(m\frac{d^2x_\nu}{dt^2} - X_\nu\right)$. These equations of motion are therefore valid in every other system of Cartesian co-ordinates in the space of reference. In the case where the forces are conservative we can easily recognize the vector character of (X_ν). For a potential energy, Φ, exists, which depends only upon the mutual distances of the particles, and is therefore an invariant. The vector character of the force, $X_\nu = -\frac{\partial\Phi}{\partial x_\nu}$, is then a consequence of our general theorem about the derivative of a tensor of rank 0.

Multiplying by the velocity, a tensor of rank 1, we obtain the tensor equation

$$\left(m\frac{d^2x_\nu}{dt^2} - X_\nu\right)\frac{dx_\mu}{dt} = 0.$$

By contraction and multiplication by the scalar dt we obtain the equation of kinetic energy

$$d\left(\frac{mq^2}{2}\right) = X_\nu dx_\nu.$$

If ξ_ν denotes the difference of the co-ordinates of the material particle and a point fixed in space, then the ξ_ν have vector character. We evidently have

$\frac{d^2x_\nu}{dt^2} = \frac{d^2\xi_\nu}{dt^2}$, so that the equations of motion of the particle may be written

$$m\frac{d^2\xi_\nu}{dt^2} - X_\nu = 0.$$

Multiplying this equation by ξ_μ we obtain a tensor equation

$$\left(m\frac{d^2\xi_\nu}{dt^2} - X_\nu\right)\xi_\mu = 0.$$

Contracting the tensor on the left and taking the time average we obtain the virial theorem, which we shall not consider further. By interchanging the indices and subsequent subtraction, we obtain, after a simple transformation, the theorem of moments,

$$(15) \qquad \frac{d}{dt}\left[m\left(\xi_\mu\frac{d\xi_\nu}{dt} - \xi_\nu\frac{d\xi_\mu}{dt}\right)\right] = \xi_\mu X_\nu - \xi_\nu X_\mu$$

It is evident in this way that the moment of a vector is not a vector but a tensor. On account of their skew-symmetrical character there are not nine, but only three independent equations of this system. The possibility of replacing skew-symmetrical tensors of the second rank in space of three dimensions by vectors depends upon the formation of the vector

$$A_\mu = \frac{1}{2}A_{\sigma\tau}\delta_{\sigma\tau\mu}.$$

If we multiply the skew-symmetrical tensor of rank 2 by the special skew-symmetrical tensor δ introduced above, and contract twice, a vector results whose components are numerically equal to those of the tensor. These are the so-called axial vectors which transform differ-

ently, from a right-handed system to a left-handed system, from the Δx_ν. There is a gain in picturesqueness in regarding a skew-symmetrical tensor of rank 2 as a vector in space of three dimensions, but it does not represent the exact nature of the corresponding quantity so well as considering it a tensor.

We consider next the equations of motion of a continuous medium. Let ρ be the density, u_ν the velocity components considered as functions of the co-ordinates and the time, X_ν the volume forces per unit of mass, and $p_{\nu\sigma}$ the stresses upon a surface perpendicular to the σ-axis in the direction of increasing x_ν. Then the equations of motion area, by Newton's law,

$$\rho \frac{du_\nu}{dt} = -\frac{\partial p_{\nu\sigma}}{\partial x_\sigma} + \rho X_\nu$$

in which $\frac{du_\nu}{dt}$ is the acceleration of the particle which at time t has the co-ordinates x_ν. If we express this acceleration by partial differential coefficients, we obtain, after dividing by ρ,

$$\frac{\partial u_\nu}{\partial t} + \frac{\partial u_\nu}{\partial x_\sigma} u_\sigma = -\frac{1}{\rho} \frac{\partial p_{\nu\sigma}}{\partial x_\sigma} + X_\nu \tag{16}$$

We must show that this equation holds independently of the special choice of the Cartesian system of co-ordinates. (u_ν) is a vector, and therefore $\frac{\partial u_\nu}{\partial t}$ is also a vector. $\frac{\partial u_\nu}{\partial x_\sigma}$ is a tensor of rank 2, $\frac{\partial u_\nu}{\partial x_\sigma} u_\tau$ is a tensor of rank 3. The second term on the left results from contraction in the indices σ, τ. The vector character of the second term on the right is obvious. In order that the first term on the right may

also be a vector it is necessary for $p_{\nu\sigma}$ to be a tensor. Then by differentiation and contraction $\frac{\partial p_{\nu\sigma}}{\partial x_\sigma}$ results, and is therefore a vector, as it also is after multiplication by the reciprocal scalar $\frac{1}{\rho}$. That $p_{\nu\sigma}$ is a tensor, and therefore transforms according to the equation

$$p'_{\mu\nu} = b_{\mu\alpha} b_{\nu\beta} p_{\alpha\beta},$$

is proved in mechanics by integrating this equation over an infinitely small tetrahedron. It is also proved there, by application of the theorem of moments to an infinitely small parallelepipedon, that $p_{\nu\sigma} = p_{\sigma\nu}$, and hence that the tensor of the stress is a symmetrical tensor. From what has been said it follows that, with the aid of the rules given above, the equation is co-variant with respect to orthogonal transformations in space (rotational transformations); and the rules according to which the quantities in the equation must be transformed in order that the equation may be co-variant also become evident.

The co-variance of the equation of continuity,

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u_\nu)}{\partial x_\nu} = 0 \tag{17}$$

requires, from the foregoing, no particular discussion.

We shall also test for co-variance the equations which express the dependence of the stress components upon the properties of the matter, and set up these equations for the case of a compressible viscous fluid with the aid of the conditions of co-variance. If we neglect the viscosity, the pressure, p, will be a scalar, and will depend only upon the density and the temperature of the fluid.

PRE-RELATIVITY PHYSICS

The contribution to the stress tensor is then evidently

$$p\delta_{\mu\nu}$$

in which $\delta_{\mu\nu}$ is the special symmetrical tensor. This term will also be present in the case of a viscous fluid. But in this case there will also be pressure terms, which depend upon the space derivatives of the u_ν. We shall assume that this dependence is a linear one. Since these terms must be symmetrical tensors, the only ones which enter will be

$$\alpha\left(\frac{\partial u_\mu}{\partial x_\nu}+\frac{\partial u_\nu}{\partial x_\mu}\right)+\beta\delta_{\mu\nu}\frac{\partial u_\alpha}{\partial x_\alpha}$$

$\left(\text{for } \frac{\partial u_\alpha}{\partial x_\alpha} \text{ is a scalar}\right)$. For physical reasons (no slipping) it is assumed that for symmetrical dilatations in all directions, i.e. when

$$\frac{\partial u_1}{\partial x_1}=\frac{\partial u_2}{\partial x_2}=\frac{\partial u_3}{\partial x_3};\ \frac{\partial u_1}{\partial x_2},\ \text{etc.},\ =0,$$

there are no frictional forces present, from which it follows that $\beta=-\frac{2}{3}\alpha$. If only $\frac{\partial u_1}{\partial x_3}$ is different from zero, let $p_{31}=-\eta\frac{\partial u_1}{\partial x_3}$, by which α is determined. We then obtain for the complete stress tensor,

$$(18)\quad p_{\mu\nu}=p\delta_{\mu\nu}-\eta\left[\left(\frac{\partial u_\mu}{\partial x_\nu}+\frac{\partial u_\nu}{\partial x_\mu}\right)-\frac{2}{3}\left(\frac{\partial u_1}{\partial x_1}+\frac{\partial u_2}{\partial x_2}+\frac{\partial u_3}{\partial x_3}\right)\delta_{\mu\nu}\right]$$

The heuristic value of the theory of invariants, which arises from the isotropy of space (equivalence of all directions), becomes evident from this example.

We consider, finally, Maxwell's equations in the form which are the foundation of the electron theory of Lorentz.

$$(19)\qquad \begin{cases} \dfrac{\partial h_3}{\partial x_2} - \dfrac{\partial h_2}{\partial x_3} = \dfrac{1}{c}\dfrac{\partial e_1}{\partial t} + \dfrac{1}{c} i_1 \\ \dfrac{\partial h_1}{\partial x_3} - \dfrac{\partial h_3}{\partial x_1} = \dfrac{1}{c}\dfrac{\partial e_2}{\partial t} + \dfrac{1}{c} i_2 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \dfrac{\partial e_1}{\partial x_1} + \dfrac{\partial e_2}{\partial x_2} + \dfrac{\partial e_3}{\partial x_3} = \rho \end{cases}$$

$$(20)\qquad \begin{cases} \dfrac{\partial e_3}{\partial x_2} - \dfrac{\partial e_2}{\partial x_3} = -\dfrac{1}{c}\dfrac{\partial h_1}{\partial t} \\ \dfrac{\partial e_1}{\partial x_3} - \dfrac{\partial e_3}{\partial x_1} = -\dfrac{1}{c}\dfrac{\partial h_2}{\partial t} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \\ \dfrac{\partial h_1}{\partial x_1} + \dfrac{\partial h_2}{\partial x_2} + \dfrac{\partial h_3}{\partial x_3} = 0 \end{cases}$$

i is a vector, because the current density is defined as the density of electricity multiplied by the vector velocity of the electricity. According to the first three equations it is evident that **e** is also to be regarded as a vector. Then **h** cannot be regarded as a vector.* The equations may, however, easily be interpreted if **h** is regarded as a skew-symmetrical tensor of the second rank. Accordingly, we write h_{23}, h_{31}, h_{12}, in place of h_1, h_2, h_3 respectively. Paying attention to the skew-symmetry of $h_{\mu\nu}$, the first three equations of (19) and (20) may be written in the form

$$(19a)\qquad \frac{\partial h_{\mu\nu}}{\partial x_\nu} = \frac{1}{c}\frac{\partial e_\mu}{\partial t} + \frac{1}{c} i_\mu$$

$$(20a)\qquad \frac{\partial e_\mu}{\partial x_\nu} - \frac{\partial e_\nu}{\partial x_\mu} = +\frac{1}{c}\frac{\partial h_{\mu\nu}}{\partial t}$$

* These considerations will make the reader familiar with tensor operations without the special difficulties of the four-dimensional treatment; corresponding considerations in the theory of special relativity (Minkowski's interpretation of the field) will then offer fewer difficulties.

PRE-RELATIVITY PHYSICS

In contrast to **e**, **h** appears as a quantity which has the same type of symmetry as an angular velocity. The divergence equations then take the form

$$\frac{\partial e_\nu}{\partial x_\nu} = \rho \tag{19b}$$

$$\frac{\partial h_{\mu\nu}}{\partial x_\rho} + \frac{\partial h_{\nu\rho}}{\partial x_\mu} + \frac{\partial h_{\rho\mu}}{\partial x_\nu} = 0 \tag{20b}$$

The last equation is a skew-symmetrical tensor equation of the third rank (the skew-symmetry of the left-hand side with respect to every pair of indices may easily be proved, if attention is paid to the skew-symmetry of $h_{\mu\nu}$). This notation is more natural than the usual one, because, in contrast to the latter, it is applicable to Cartesian left-handed systems as well as to right-handed systems without change of sign.

THE THEORY OF SPECIAL RELATIVITY

THE previous considerations concerning the configuration of rigid bodies have been founded, irrespective of the assumption as to the validity of the Euclidean geometry, upon the hypothesis that all directions in space, or all configurations of Cartesian systems of co-ordinates, are physically equivalent. We may express this as the "principle of relativity with respect to direction," and it has been shown how equations (laws of nature) may be found, in accord with this principle, by the aid of the calculus of tensors. We now inquire whether there is a relativity with respect to the state of motion of the space of reference; in other words, whether there are spaces of reference in motion relatively to each other which are physically equivalent. From the standpoint of mechanics it appears that equivalent spaces of reference do exist. For experiments upon the earth tell us nothing of the fact that we are moving about the sun with a velocity of approximately 30 kilometres a second. On the other hand, this physical equivalence does not seem to hold for spaces of reference in arbitrary motion; for mechanical effects do not seem to be subject to the same laws in a jolting railway train as in one moving with uniform velocity; the rotation of the earth must be considered in writing down the equations of motion relatively to the earth. It appears, therefore, as if there were Cartesian systems of co-ordinates, the so-called inertial systems, with reference

to which the laws of mechanics (more generally the laws of physics) are expressed in the simplest form. We may surmise the validity of the following proposition: If K is an inertial system, then every other system K' which moves uniformly and without rotation relatively to K, is also an inertial system; the laws of nature are in concordance for all inertial systems. This statement we shall call the "principle of special relativity." We shall draw certain conclusions from this principle of "relativity of translation" just as we have already done for relativity of direction.

In order to be able to do this, we must first solve the following problem. If we are given the Cartesian co-ordinates, x, and the time t, of an event relatively to one inertial system, K, how can we calculate the co-ordinates, x', and the time, t', of the same event relatively to an inertial system K' which moves with uniform translation relatively to K? In the pre-relativity physics this problem was solved by making unconsciously two hypotheses:—

1. Time is absolute; the time of an event, t', relatively to K' is the same as the time relatively to K. If instantaneous signals could be sent to a distance, and if one knew that the state of motion of a clock had no influence on its rate, then this assumption would be physically validated. For then clocks, similar to one another, and regulated alike, could be distributed over the systems K and K', at rest relatively to them, and their indications would be independent of the state of motion of the systems; the time of an event would then be given by the clock in its immediate neighbourhood.

2. Length is absolute; if an interval, at rest relatively to K, has a length s, then it has the same length s, relatively to a system K' which is in motion relatively to K.

SPECIAL RELATIVITY

If the axes of K and K' are parallel to each other, a simple calculation based on these two assumptions, gives the equations of transformation

$$(21) \qquad \begin{cases} x'_\nu = x_\nu - a_\nu - b_\nu t \\ t' = t - b \end{cases}$$

This transformation is known as the "Galilean Transformation." Differentiating twice by the time, we get

$$\frac{d^2x'_\nu}{dt^2} = \frac{d^2x_\nu}{dt^2}.$$

Further, it follows that for two simultaneous events,

$$x'_\nu{}^{(1)} - x'_\nu{}^{(2)} = x_\nu{}^{(1)} - x_\nu{}^{(2)}.$$

The invariance of the distance between the two points results from squaring and adding. From this easily follows the co-variance of Newton's equations of motion with respect to the Galilean transformation (21). Hence it follows that classical mechanics is in accord with the principle of special relativity if the two hypotheses respecting scales and clocks are made.

But this attempt to found relativity of translation upon the Galilean transformation fails when applied to electromagnetic phenomena. The Maxwell-Lorentz electromagnetic equations are not co-variant with respect to the Galilean transformation. In particular, we note, by (21), that a ray of light which referred to K has a velocity c, has a different velocity referred to K', depending upon its direction. The space of reference of K is therefore distinguished, with respect to its physical properties, from all spaces of reference which are in motion relatively to it (quiescent ether). But all experiments have shown that electro-magnetic and optical phenomena, relatively to the

earth as the body of reference, are not influenced by the translational velocity of the earth. The most important of these experiments are those of Michelson and Morley, which I shall assume are known. The validity of the principle of special relativity also with respect to electromagnetic phenomena can therefore hardly be doubted.

On the other hand, the Maxwell-Lorentz equations have proved their validity in the treatment of optical problems in moving bodies. No other theory has satisfactorily explained the facts of aberration, the propagation of light in moving bodies (Fizeau), and phenomena observed in double stars (De Sitter). The consequence of the Maxwell-Lorentz equations that in a vacuum light is propagated with the velocity c, at least with respect to a definite inertial system K, must therefore be regarded as proved. According to the principle of special relativity, we must also assume the truth of this principle for every other inertial system.

Before we draw any conclusions from these two principles we must first review the physical significance of the concepts "time" and "velocity." It follows from what has gone before, that co-ordinates with respect to an inertial system are physically defined by means of measurements and constructions with the aid of rigid bodies. In order to measure time, we have supposed a clock, U, present somewhere, at rest relatively to K. But we cannot fix the time, by means of this clock, of an event whose distance from the clock is not negligible; for there are no "instantaneous signals" that we can use in order to compare the time of the event with that of the clock. In order to complete the definition of time we may employ the principle of the constancy of the velocity of light in a vacuum. Let us suppose that we place similar clocks at points of the

system K, at rest relatively to it, and regulated according to the following scheme. A ray of light is sent out from one of the clocks, U_m, at the instant when it indicates the time t_m, and travels through a vacuum a distance r_{mn}, to the clock U_n; at the instant when this ray meets the clock U_n the latter is set to indicate the time $t_n = t_m + \frac{r_{mn}}{c}$.* The principle of the constancy of the velocity of light then states that this adjustment of the clocks will not lead to contradictions. With clocks so adjusted, we can assign the time to events which take place near any one of them. It is essential to note that this definition of time relates only to the inertial system K, since we have used a system of clocks at rest relatively to K. The assumption which was made in the pre-relativity physics of the absolute character of time (i.e. the independence of time of the choice of the inertial system) does not follow at all from this definition.

The theory of relativity is often criticized for giving, without justification, a central theoretical rôle to the propagation of light, in that it founds the concept of time upon the law of propagation of light. The situation, however, is somewhat as follows. In order to give physical significance to the concept of time, processes of some kind are required which enable relations to be established between different places. It is immaterial what kind of processes one chooses for such a definition of time. It is advantageous, however, for the theory, to choose only those processes concerning which we know something

* Strictly speaking, it would be more correct to define simultaneity first, somewhat as follows: two events taking place at the points A and B of the system K are simultaneous if they appear at the same instant when observed from the middle point, M, of the interval AB. Time is then defined as the ensemble of the indications of similar clocks, at rest relatively to K, which register the same time simultaneously.

certain. This holds for the propagation of light *in vacuo* in a higher degree than for any other process which could be considered, thanks to the investigations of Maxwell and H. A. Lorentz.

From all of these considerations, space and time data have a physically real, and not a mere fictitious, significance; in particular this holds for all the relations in which co-ordinates and time enter, e.g. the relations (21). There is, therefore, sense in asking whether those equations are true or not, as well as in asking what the true equations of transformation are by which we pass from one inertial system K to another, K', moving relatively to it. It may be shown that this is uniquely settled by means of the principle of the constancy of the velocity of light and the principle of special relativity.

To this end we think of space and time physically defined with respect to two inertial systems, K and K', in the way that has been shown. Further, let a ray of light pass from one point P_1 to another point P_2 of K through a vacuum. If r is the measured distance between the two points, then the propagation of light must satisfy the equation

$$r = c \cdot \Delta t.$$

If we square this equation, and express r^2 by the differences of the co-ordinates, Δx_ν, in place of this equation we can write

$$\sum (\Delta x_\nu)^2 - c^2 \Delta t^2 = 0 \tag{22}$$

This equation formulates the principle of the constancy of the velocity of light relatively to K. It must hold whatever may be the motion of the source which emits the ray of light.

The same propagation of light may also be considered relatively to K', in which case also the principle of the constancy of the velocity of light must be satisfied. Therefore, with respect to K', we have the equation

$$\sum (\Delta x'_\nu)^2 - c^2 \Delta t'^2 = 0 \tag{22a}$$

Equations (22a) and (22) must be mutually consistent with each other with respect to the transformation which transforms from K to K'. A transformation which effects this we shall call a "Lorentz transformation."

Before considering these transformations in detail we shall make a few general remarks about space and time. In the pre-relativity physics space and time were separate entities. Specifications of time were independent of the choice of the space of reference. The Newtonian mechanics was relative with respect to the space of reference, so that, e.g. the statement that two non-simultaneous events happened at the same place had no objective meaning (that is, independent of the space of reference). But this relativity had no rôle in building up the theory. One spoke of points of space, as of instants of time, as if they were absolute realities. It was not observed that the true element of the space-time specification was the event specified by the four numbers x_1, x_2, x_3, t. The conception of something happening was always that of a four-dimensional continuum; but the recognition of this was obscured by the absolute character of the pre-relativity time. Upon giving up the hypothesis of the absolute character of time, particularly that of simultaneity, the four-dimensionality of the time-space concept was immediately recognized. It is neither the point in space, nor the instant in time, at which something happens that has physical reality, but only the event itself. There is no absolute (independent of the space

of reference) relation in space, and no absolute relation in time between two events, but there is an absolute (independent of the space of reference) relation in space and time, as will appear in the sequel. The circumstance that there is no objective rational division of the four-dimensional continuum into a three-dimensional space and a one-dimensional time continuum indicates that the laws of nature will assume a form which is logically most satisfactory when expressed as laws in the four-dimensional space-time continuum. Upon this depends the great advance in method which the theory of relativity owes to Minkowski. Considered from this standpoint, we must regard x_1, x_2, x_3, t as the four co-ordinates of an event in the four-dimensional continuum. We have far less success in picturing to ourselves relations in this four-dimensional continuum than in the three-dimensional Euclidean continuum; but it must be emphasized that even in the Euclidean three-dimensional geometry its concepts and relations are only of an abstract nature in our minds, and are not at all identical with the images we form visually and through our sense of touch. The non-divisibility of the four-dimensional continuum of events does not at all, however, involve the equivalence of the space co-ordinates with the time co-ordinate. On the contrary, we must remember that the time co-ordinate is defined physically wholly differently from the space co-ordinates. The relations (22) and (22a) which when equated define the Lorentz transformation show, further, a difference in the rôle of the time co-ordinate from that of the space co-ordinates; for the term Δt^2 has the opposite sign to the space terms, Δx_1^2, Δx_2^2, Δx_3^2.

Before we analyse further the conditions which define the Lorentz transformation, we shall introduce the light-time, $l = ct$, in place of the time, t, in order that the con-

stant c shall not enter explicitly into the formulas to be developed later. Then the Lorentz transformation is defined in such a way that, first, it makes the equation

$$\Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 - \Delta l^2 = 0 \tag{22b}$$

a co-variant equation, that is, an equation which is satisfied with respect to every inertial system if it is satisfied in the inertial system to which we refer the two given events (emission and reception of the ray of light). Finally, with Minkowski, we introduce in place of the real time co-ordinate $l = ct$, the imaginary time co-ordinate

$$x_4 = il = ict \ (\sqrt{-1} = i).$$

Then the equation defining the propagation of light, which must be co-variant with respect to the Lorentz transformation, becomes

$$\sum_{(4)} \Delta x_\nu^2 = \Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 + \Delta x_4^2 = 0 \tag{22c}$$

This condition is always satisfied* if we satisfy the more general condition that

$$s^2 = \Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 + \Delta x_4^2 \tag{23}$$

shall be an invariant with respect to the transformation. This condition is satisfied only by linear transformations, that is, transformations of the type

$$x'_\mu = a_\mu + b_{\mu\alpha} x_\alpha \tag{24}$$

in which the summation over the α is to be extended from $\alpha = 1$ to $\alpha = 4$. A glance at equations (23) (24) shows that the Lorentz transformation so defined is identical

* That this specialization lies in the nature of the case will be evident later.

with the translational and rotational transformations of the Euclidean geometry, if we disregard the number of dimensions and the relations of reality. We can also conclude that the coefficients $b_{\mu\alpha}$ must satisfy the conditions

$$b_{\mu\alpha}b_{\nu\alpha} = \delta_{\mu\nu} = b_{\alpha\mu}b_{\alpha\nu} \tag{25}$$

Since the ratios of the x_ν are real, it follows that all the a_μ and the $b_{\mu\alpha}$ are real, except a_4, b_{41}, b_{42}, b_{43}, b_{14}, b_{24}, and b_{34}, which are purely imaginary.

Special Lorentz Transformation. We obtain the simplest transformations of the type of (24) and (25) if only two of the co-ordinates are to be transformed, and if all the a_μ, which merely determine the new origin, vanish. We obtain then for the indices 1 and 2, on account of the three independent conditions which the relations (25) furnish,

$$\begin{cases} x'_1 = x_1 \cos\phi - x_2 \sin\phi \\ x'_2 = x_1 \sin\phi + x_2 \cos\phi \\ x'_3 = x_3 \\ x'_4 = x_4 \end{cases} \tag{26}$$

This is a simple rotation in space of the (space) co-ordinate system about the x_3-axis. We see that the rotational transformation in space (without the time transformation) which we studied before is contained in the Lorentz transformation as a special case. For the indices 1 and 4 we obtain, in an analogous manner,

$$\begin{cases} x'_1 = x_1 \cos\psi - x_4 \sin\psi \\ x'_4 = x_1 \sin\psi + x_4 \cos\psi \\ x'_2 = x_2 \\ x'_3 = x_3 \end{cases} \tag{26a}$$

On account of the relations of reality ψ must be taken as imaginary. To interpret these equations physically,

we introduce the real light-time l and the velocity v of K' relatively to K, instead of the imaginary angle ψ. We have, first,

$$x'_1 = x_1 \cos\psi - il \sin\psi$$
$$l' = -ix_1 \sin\psi + l \cos\psi.$$

Since for the origin of K', i.e., for $x'_1 = 0$, we must have $x_1 = vl$, it follows from the first of these equations that

(27) $$v = i \tan\psi$$

and also

(28) $$\begin{cases} \sin\psi = \dfrac{-iv}{\sqrt{1-v^2}} \\ \cos\psi = \dfrac{1}{\sqrt{1-v^2}} \end{cases}$$

so that we obtain

(29) $$\begin{cases} x'_1 = \dfrac{x_1 - vl}{\sqrt{1-v^2}} \\ l' = \dfrac{l - vx_1}{\sqrt{1-v^2}} \\ x'_2 = x_2 \\ x'_3 = x_3 \end{cases}$$

These equations form the well-known special Lorentz transformation, which in the general theory represents a rotation, through an imaginary angle, of the four-dimensional system of co-ordinates. If we introduce the ordinary time t, in place of the light-time l, then in (29) we must replace l by ct and v by $\frac{v}{c}$.

We must now fill in a gap. From the principle of the constancy of the velocity of light it follows that the equation

$$\sum \Delta x_\nu^2 = 0$$

has a significance which is independent of the choice of the inertial system; but the invariance of the quantity $\sum \Delta x_\nu^2$ does not at all follow from this. This quantity might be transformed with a factor. This depends upon the fact that the right-hand side of (29) might be multiplied by a factor λ, which may depend on v. But the principle of relativity does not permit this factor to be different from 1, as we shall now show. Let us assume that we have a rigid circular cylinder moving in the direction of its axis. If its radius, measured at rest with a unit measuring rod is equal to R_0, its radius R in motion, might be different from R_0, since the theory of relativity does not make the assumption that the shape of bodies with respect to a space of reference is independent of their motion relatively to this space of reference. But all directions in space must be equivalent to each other. R may therefore depend upon the magnitude q of the velocity, but not upon its direction; R must therefore be an even function of q. If the cylinder is at rest relatively to K' the equation of its lateral surface is

$$x'^2 + y'^2 = R_0^2.$$

If we write the last two equations of (29) more generally

$$x'_2 = \lambda x_2$$
$$x'_3 = \lambda x_3$$

then the lateral surface of the cylinder referred to K satisfies the equation

$$x^2 + y^2 = \frac{R_0^2}{\lambda^2}.$$

The factor λ therefore measures the lateral contraction of

the cylinder, and can thus, from the above, be only an even function of v.

If we introduce a third system of co-ordinates, K'', which moves relatively to K' with velocity v in the direction of the negative x-axis of K, we obtain, by applying (29) twice,

$$x''_1 = \lambda(v)\lambda(-v)x_1$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot$$
$$l'' = \lambda(v)\lambda(-v)l.$$

Now, since $\lambda(v)$ must be equal to $\lambda(-v)$, and since we assume that we use the same measuring rods in all the systems, it follows that the transformation of K'' to K must be the identical transformation (since the possibility $\lambda = -1$ does not need to be considered). It is essential for these considerations to assume that the behaviour of the measuring rods does not depend upon the history of their previous motion.

Moving Measuring Rods and Clocks. At the definite K time, $l = 0$, the position of the points given by the integers $x'_1 = n$, is with respect to K, given by $x_1 = n\sqrt{1 - v^2}$; this follows from the first of equations (29) and expresses the Lorentz contraction. A clock at rest at the origin $x_1 = 0$ of K, whose beats are characterized by $l = n$, will, when observed from K', have beats characterized by

$$l' = \frac{n}{\sqrt{1 - v^2}};$$

this follows from the second of equations (29) and shows that the clock goes slower than if it were at rest relatively to K'. These two consequences, which hold, *mutatis mutandis*, for

every system of reference, form the physical content, free from convention, of the Lorentz transformation.

Addition Theorem for Velocities. If we combine two special Lorentz transformations with the relative velocities v_1 and v_2, then the velocity of the single Lorentz transformation which takes the place of the two separate ones is, according to (27), given by

$$(30) \quad v_{12} = i \tan(\psi_1 + \psi_2) = i \frac{\tan\psi_1 + \tan\psi_2}{1 - \tan\psi_1 \tan\psi_2} = \frac{v_1 + v_2}{1 + v_1 v_2}.$$

General Statements about the Lorentz Transformation and its Theory of Invariants. The whole theory of invariants of the special theory of relativity depends upon the invariant s^2 (23). Formally, it has the same rôle in the four-dimensional space-time continuum as the invariant $\Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2$ in the Euclidean geometry and in the pre-relativity physics. The latter quantity is not an invariant with respect to all the Lorentz transformations; the quantity s^2 of equation (23) assumes the rôle of this invariant. With respect to an arbitrary inertial system, s^2 may be determined by measurements; with a given unit of measure it is a completely determinate quantity, associated with an arbitrary pair of events.

The invariant s^2 differs, disregarding the number of dimensions, from the corresponding invariant of the Euclidean geometry in the following points. In the Euclidean geometry s^2 is necessarily positive; it vanishes only when the two points concerned come together. On the other hand, from the vanishing of

$$s^2 = \sum \Delta x_\nu^2 = \Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 - \Delta t^2$$

it cannot be concluded that the two space-time points

fall together; the vanishing of this quantity s^2, is the invariant condition that the two space-time points can be connected by a light signal *in vacuo*. If P is a point (event) represented in the four-dimensional space of the x_1, x_2, x_3, l, then all the "points" which can be connected to P by means of a light signal lie upon the cone $s^2 = 0$ (compare Fig. 1, in which the dimension x_3 is suppressed). The "upper"

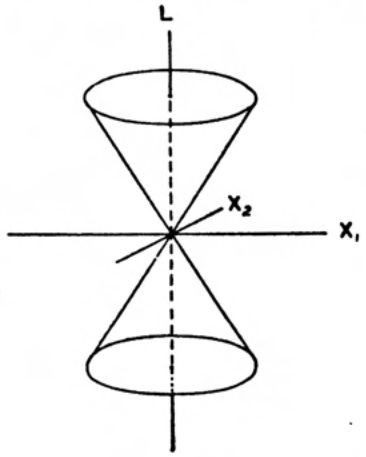


FIG. 1.

half of the cone may contain the "points" to which light signals can be sent from P; then the "lower" half of the cone will contain the "points" from which light signals can be sent to P. The points P' enclosed by the conical surface furnish, with P, a negative s^2; PP', as well as $P'P$ is then, according to Minkowski, time-like. Such intervals represent elements of possible paths of motion, the velocity being less than that of light.* In this case the l-axis

* That material velocities exceeding that of light are not possible, follows from the appearance of the radical $\sqrt{1 - v^2}$ in the special Lorentz transformation (29).

may be drawn in the direction of PP' by suitably choosing the state of motion of the inertial system. If P' lies outside of the "light-cone" then PP' is space-like; in this case, by properly choosing the inertial system, Δl can be made to vanish.

By the introduction of the imaginary time variable, $x_4 = il$, Minkowski has made the theory of invariants for the four-dimensional continuum of physical phenomena fully analogous to the theory of invariants for the three-dimensional continuum of Euclidean space. The theory of four-dimensional tensors of special relativity differs from the theory of tensors in three-dimensional space, therefore, only in the number of dimensions and the relations of reality.

A physical entity which is specified by four quantities, A_ν, in an arbitrary inertial system of the x_1, x_2, x_3, x_4, is called a 4-vector, with the components A_ν, if the A_ν correspond in their relations of reality and the properties of transformation to the Δx_ν; it may be space-like or time-like. The sixteen quantities $A_{\mu\nu}$ then form the components of a tensor of the second rank, if they transform according to the scheme

$$A'_{\mu\nu} = b_{\mu\alpha} b_{\nu\beta} A_{\alpha\beta}.$$

It follows from this that the $A_{\mu\nu}$ behave, with respect to their properties of transformation and their properties of reality, as the products of the components, U_μ, V_ν, of two 4-vectors, (U) and (V). All the components are real except those which contain the index 4 once, those being purely imaginary. Tensors of the third and higher ranks may be defined in an analogous way. The operations of addition, subtraction, multiplication, contraction and differentiation for these tensors are wholly analogous to the corresponding operations for tensors in three-dimensional space.

SPECIAL RELATIVITY

Before we apply the tensor theory to the four-dimensional space-time continuum, we shall examine more particularly the skew-symmetrical tensors. The tensor of the second rank has, in general, $16 = 4.4$ components. In the case of skew-symmetry the components with two equal indices vanish, and the components with unequal indices are equal and opposite in pairs. There exist, therefore, only six independent components, as is the case in the electromagnetic field. In fact, it will be shown when we consider Maxwell's equations that these may be looked upon as tensor equations, provided we regard the electromagnetic field as a skew-symmetrical tensor. Further, it is clear that a skew-symmetrical tensor of the third rank (skew-symmetrical in all pairs of indices) has only four independent components, since there are only four combinations of three different indices.

We now turn to Maxwell's equations (19a), (19b), (20a), (20b), and introduce the notation:*

$$(30a)\qquad \begin{cases} \phi_{23} & \phi_{31} & \phi_{12} & \phi_{14} & \phi_{24} & \phi_{34} \\ h_{23} & h_{31} & h_{12} & -ie_x & -ie_y & -ie_z \end{cases}$$

$$(31)\qquad \begin{cases} \mathcal{J}_1 & \mathcal{J}_2 & \mathcal{J}_3 & \mathcal{J}_4 \\ \frac{1}{c}i_x & \frac{1}{c}i_y & \frac{1}{c}i_z & i\rho \end{cases}$$

with the convention that $\phi_{\mu\nu}$ shall be equal to $-\phi_{\nu\mu}$. Then Maxwell's equations may be combined into the forms

$$(32)\qquad \frac{\partial \phi_{\mu\nu}}{\partial x_\nu} = \mathcal{J}_\mu$$

$$(33)\qquad \frac{\partial \phi_{\mu\nu}}{\partial x_\sigma} + \frac{\partial \phi_{\nu\sigma}}{\partial x_\mu} + \frac{\partial \phi_{\sigma\mu}}{\partial x_\nu} = 0$$

* In order to avoid confusion from now on we shall use the three-dimensional space indices, x, y, z instead of 1, 2, 3, and we shall reserve the numeral indices 1, 2, 3, 4 for the four-dimensional space-time continuum.

as one can easily verify by substituting from (30a) and (31). Equations (32) and (33) have a tensor character, and are therefore co-variant with respect to Lorentz transformations, if the $\phi_{\mu\nu}$ and the $\mathfrak{J}_\mu$ have a tensor character, which we assume. Consequently, the laws for transforming these quantities from one to another allowable (inertial) system of co-ordinates are uniquely determined. The progress in method which electro-dynamics owes to the theory of special relativity lies principally in this, that the number of independent hypotheses is diminished. If we consider, for example, equations (19a) only from the standpoint of relativity of direction, as we have done above, we see that they have three logically independent terms. The way in which the electric intensity enters these equations appears to be wholly independent of the way in which the magnetic intensity enters them; it would not be surprising if instead of $\frac{\partial e_\mu}{\partial l}$, we had, say, $\frac{\partial^2 e_\mu}{\partial l^2}$, or if this term were absent. On the other hand, only two independent terms appear in equation (32). The electromagnetic field appears as a formal unit; the way in which the electric field enters this equation is determined by the way in which the magnetic field enters it. Besides the electromagnetic field, only the electric current density appears as an independent entity. This advance in method arises from the fact that the electric and magnetic fields lose their separate existences through the relativity of motion. A field which appears to be purely an electric field, judged from one system, has also magnetic field components when judged from another inertial system. When applied to an electromagnetic field, the general law of transformation furnishes, for the special case of the special Lorentz transformation, the equations

$$(34)\qquad \begin{cases} e'_x = e_x & h'_x = h_x \\ e'_y = \dfrac{e_y - v h_z}{\sqrt{1 - v^2}} & h'_y = \dfrac{h_y + v e_z}{\sqrt{1 - v^2}} \\ e'_z = \dfrac{e_z + v h_y}{\sqrt{1 - v^2}} & h'_z = \dfrac{h_z - v e_y}{\sqrt{1 - v^2}} \end{cases}$$

If there exists with respect to K only a magnetic field, **h,** but no electric field, **e,** then with respect to K' there exists an electric field $\mathbf{e}'$ as well, which would act upon an electric particle at rest relatively to K'. An observer at rest relatively to K would designate this force as the Biot-Savart force, or the Lorentz electromotive force. It therefore appears as if this electromotive force had become fused with the electric field intensity into a single entity.

In order to view this relation formally, let us consider the expression for the force acting upon unit volume of electricity,

$$(35)\qquad \mathbf{k} = \rho\mathbf{e} + \mathbf{i} \times \mathbf{h}$$

in which **i** is the vector velocity of electricity, with the velocity of light as the unit. If we introduce $\mathcal{J}_\mu$ and ϕ_μ according to (30a) and (31), we obtain for the first component the expression

$$\phi_{12}\mathcal{J}_2 + \phi_{13}\mathcal{J}_3 + \phi_{14}\mathcal{J}_4.$$

Observing that ϕ_{11} vanishes on account of the skew-symmetry of the tensor (ϕ), the components of k are given by the first three components of the four-dimensional vector

$$(36)\qquad K_\mu = \phi_{\mu\nu}\mathcal{J}_\nu$$

and the fourth component is given by

$$(37)\qquad K_4 = \phi_{41}\mathcal{J}_1 + \phi_{42}\mathcal{J}_2 + \phi_{43}\mathcal{J}_3 = i(e_x i_x + e_y i_y + e_z i_z) = i\lambda.$$

There is, therefore, a four-dimensional vector of force per unit volume, whose first three components, k_1, k_2, k_3, are the ponderomotive force components per unit volume, and whose fourth component is the rate of working of the field per unit volume, multiplied by $\sqrt{-1}$.

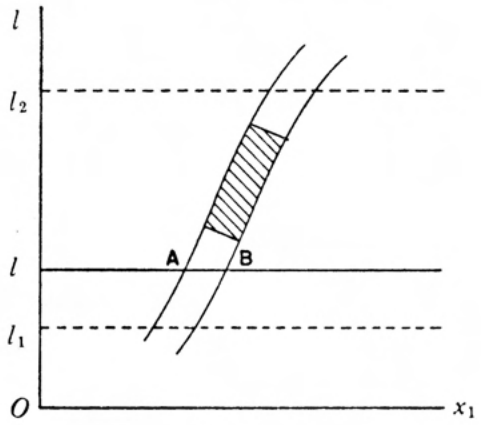


FIG. 2.

A comparison of (36) and (35) shows that the theory of relativity formally unites the ponderomotive force of the electric field, $\rho\mathbf{e}$, and the Biot-Savart or Lorentz force $\mathbf{i} \times \mathbf{h}$.

Mass and Energy. An important conclusion can be drawn from the existence and significance of the 4-vector K_μ. Let us imagine a body upon which the electromagnetic field acts for a time. In the symbolic figure (Fig. 2) Ox_1 designates the x_1-axis, and is at the same time a substitute for the three space axes Ox_1, Ox_2, Ox_3; Ol designates the real time axis. In this diagram a body of finite extent is represented, at a definite time l, by the interval AB; the whole space-time existence of the body is represented by a strip whose boundary is everywhere

inclined less than 45° to the l-axis. Between the time sections, $l = l_1$ and $l = l_2$, but not extending to them, a portion of the strip is shaded. This represents the portion of the space-time manifold in which the electromagnetic field acts upon the body, or upon the electric charges contained in it, the action upon them being transmitted to the body. We shall now consider the changes which take place in the momentum and energy of the body as a result of this action.

We shall assume that the principles of momentum and energy are valid for the body. The change in momentum, ΔI_x, ΔI_y, ΔI_z, and the change in energy, ΔE, are then given by the expressions

$$\Delta I_x = \int_{l_0}^{l_1} dl \int k_x dx dy dz = \frac{1}{i} \int K_1 dx_1 dx_2 dx_3 dx_4$$

$$\cdots \quad \cdots$$

$$\Delta E = \int_{l_0}^{l_1} dl \int \lambda dx dy dz = \frac{1}{i} \int \frac{1}{i} K_4 dx_1 dx_2 dx_3 dx_4$$

Since the four-dimensional element of volume is an invariant, and (K_1, K_2, K_3, K_4) forms a 4-vector, the four-dimensional integral extended over the shaded portion transforms as a 4-vector, as does also the integral between the limits l_1 and l_2, because the portion of the region which is not shaded contributes nothing to the integral. It follows, therefore, that ΔI_x, ΔI_y, ΔI_z, $i\Delta E$ form a 4-vector. Since the quantities themselves may be presumed to transform in the same way as their increments, we infer that the aggregate of the four quantities

$$I_x, I_y, I_z, iE$$

has itself vector character; these quantities are referred

to an instantaneous condition of the body (e.g. at the time $l = l_1$).

This 4-vector may also be expressed in terms of the mass m, and the velocity of the body, considered as a material particle. To form this expression, we note first, that

$$-ds^2 = d\tau^2 = -(dx_1^2 + dx_2^2 + dx_3^2) - dx_4^2 = dl^2(1 - q^2) \tag{38}$$

is an invariant which refers to an infinitely short portion of the four-dimensional line which represents the motion of the material particle. The physical significance of the invariant $d\tau$ may easily be given. If the time axis is chosen in such a way that it has the direction of the line differential which we are considering, or, in other terms, if we transform the material particle to rest, we shall have $d\tau = dl$; this will therefore be measured by the light-seconds clock which is at the same place, and at rest relatively to the material particle. We therefore call τ the proper time of the material particle. As opposed to dl, $d\tau$ is therefore an invariant, and is practically equivalent to dl for motions whose velocity is small compared to that of light. Hence we see that

$$u_\sigma = \frac{dx_\sigma}{d\tau} \tag{39}$$

has, just as the dx_ν, the character of a vector; we shall designate (u_σ) as the four-dimensional vector (in brief, 4-vector) of velocity. Its components satisfy, by (38), the condition

$$\sum u_\sigma^2 = -1. \tag{40}$$

We see that this 4-vector, whose components in the ordinary notation are

(41) $$\frac{q_x}{\sqrt{1-q^2}}, \frac{q_y}{\sqrt{1-q^2}}, \frac{q_z}{\sqrt{1-q^2}}, \frac{i}{\sqrt{1-q^2}}$$

is the only 4-vector which can be formed from the velocity components of the material particle which are defined in three dimensions by

$$q_x = \frac{dx}{dl}, \quad q_y = \frac{dy}{dl}, \quad q_z = \frac{dz}{dl}.$$

We therefore see that

(42) $$\left(m\frac{dx_\mu}{d\tau}\right)$$

must be that 4-vector which is to be equated to the 4-vector of momentum and energy whose existence we have proved above. By equating the components, we obtain, in three-dimensional notation,

(43) $$\begin{cases} I_x = \dfrac{mq_x}{\sqrt{1-q^2}} \\ \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \\ E = \dfrac{m}{\sqrt{1-q^2}} \end{cases}$$

We recognize, in fact, that these components of momentum agree with those of classical mechanics for velocities which are small compared to that of light. For large velocities the momentum increases more rapidly than linearly with the velocity, so as to become infinite on approaching the velocity of light.

If we apply the last of equations (43) to a material particle at rest ($q = 0$), we see that the energy, E_0, of a body at rest is equal to its mass. Had we chosen the second as our unit of time, we would have obtained

(44) $$E_0 = mc^2.$$

Mass and energy are therefore essentially alike; they are only different expressions for the same thing. The mass of a body is not a constant; it varies with changes in its energy.* We see from the last of equations (43) that E becomes infinite when q approaches 1, the velocity of light. If we develop E in powers of q^2, we obtain,

$$E = m + \frac{m}{2} q^2 + \frac{3}{8} mq^4 + \ldots \tag{45}$$

The second term of this expansion corresponds to the kinetic energy of the material particle in classical mechanics.

Equations of Motion of Material Particles. From (43) we obtain, by differentiating by the time l, and using the principle of momentum, in the notation of three-dimensional vectors,

$$\mathbf{K} = \frac{d}{dl}\left(\frac{m\mathbf{q}}{\sqrt{1 - q^2}}\right) \tag{46}$$

This equation, which was previously employed by H. A. Lorentz for the motion of electrons, has been proved to be true, with great accuracy, by experiments with β-rays.

Energy Tensor of the Electromagnetic Field. Before the development of the theory of relativity it was known that the principles of energy and momentum could be expressed in a differential form for the electromagnetic field. The four-dimensional formulation of these principles leads to an important conception, that of the energy tensor, which is important for the further development of the theory of relativity.

* The emission of energy in radioactive processes is evidently connected with the fact that the atomic weights are not integers. The equivalence between mass at rest and energy at rest which is expressed in equation (44) has been confirmed in many cases during recent years. In radio-active decomposition the sum of the resulting masses is always less than the mass of the decomposing atom. The difference appears in the form of kinetic energy of the generated particles as well as in the form of released radiational energy.

SPECIAL RELATIVITY

If in the expression for the 4-vector of force per unit volume,

$$K_\mu = \phi_{\mu\nu}\mathcal{J}_\nu$$

using the field equations (32), we express $\mathcal{J}_\mu$ in terms of the field intensities, $\phi_{\mu\nu}$, we obtain, after some transformations and repeated application of the field equations (32) and (33), the expression

$$K_\mu = -\frac{\partial T_{\mu\nu}}{\partial x_\nu} \tag{47}$$

where we have written*

$$T_{\mu\nu} = -\tfrac{1}{4}\phi_{\alpha\beta}{}^2\delta_{\mu\nu} + \phi_{\mu\alpha}\phi_{\nu\alpha} \tag{48}$$

The physical meaning of equation (47) becomes evident if in place of this equation we write, using a new notation,

$$\text{(47a)}\qquad \begin{cases} k_x = -\dfrac{\partial p_{xx}}{\partial x} - \dfrac{\partial p_{xy}}{\partial y} - \dfrac{\partial p_{xz}}{\partial z} - \dfrac{\partial (ib_x)}{\partial (il)} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ i\lambda = -\dfrac{\partial (is_x)}{\partial x} - \dfrac{\partial (is_y)}{\partial y} - \dfrac{\partial (is_z)}{\partial z} - \dfrac{\partial (-\eta)}{\partial (il)} \end{cases}$$

or, on eliminating the imaginary,

$$\text{(47b)}\qquad \begin{cases} k_x = -\dfrac{\partial p_{xx}}{\partial x} - \dfrac{\partial p_{xy}}{\partial y} - \dfrac{\partial p_{xz}}{\partial z} - \dfrac{\partial b_x}{\partial l} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \lambda = -\dfrac{\partial s_x}{\partial x} - \dfrac{\partial s_y}{\partial y} - \dfrac{\partial s_z}{\partial z} - \dfrac{\partial \eta}{\partial l} \end{cases}$$

When expressed in the latter form, we see that the first three equations state the principle of momentum; p_{xx} . . . p_{zz} are the Maxwell stresses in the electro-magnetic

* To be summed for the indices α and β.

field, and (b_x, b_y, b_z) is the vector momentum per unit volume of the field. The last of equations (47b) expresses the energy principle; **s** is the vector flow of energy, and η the energy per unit volume of the field. In fact, we get from (48) by introducing the real components of the field intensity the following expressions well known from electrodynamics:

$$
\text{(48a)}\quad
\begin{cases}
p_{xx} = -h_x h_x + \frac{1}{2}(h_x^2 + h_y^2 + h_z^2) \\
\qquad -e_x e_x + \frac{1}{2}(e_x^2 + e_y^2 + e_z^2) \\
\qquad\qquad\qquad\qquad p_{xy} = -h_x h_y - e_x e_y \\
\qquad\qquad\qquad\qquad p_{xz} = -h_x h_z - e_x e_z \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
b_x = s_x = e_y h_z - e_z h_y \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\eta = +\frac{1}{2}(e_x^2 + e_y^2 + e_z^2 + h_x^2 + h_y^2 + h_z^2)
\end{cases}
$$

We notice from (48) that the energy tensor of the electromagnetic field is symmetrical; with this is connected the fact that the momentum per unit volume and the flow of energy are equal to each other (relation between energy and inertia).

We therefore conclude from these considerations that the energy per unit volume has the character of a tensor. This has been proved directly only for an electromagnetic field, although we may claim universal validity for it. Maxwell's equations determine the electromagnetic field when the distribution of electric charges and currents is known. But we do not know the laws which govern the currents and charges. We do know, indeed, that electricity consists of elementary particles (electrons, positive nuclei), but from a theoretical point of view we cannot comprehend this. We do not know the energy factors which determine

the distribution of electricity in particles of definite size and charge, and all attempts to complete the theory in this direction have failed. If then we can build upon Maxwell's equations at all, the energy tensor of the electromagnetic field is known only outside the charged particles.* In these regions, outside of charged particles, the only regions in which we can believe that we have the complete expression for the energy tensor, we have, by (47),

$$\frac{\partial T_{\mu\nu}}{\partial x_\nu} = 0. \tag{47c}$$

General Expressions for the Conservation Principles. We can hardly avoid making the assumption that in all other cases, also, the space distribution of energy is given by a symmetrical tensor, $T_{\mu\nu}$, and that this complete energy tensor everywhere satisfies the relation (47c). At any rate we shall see that by means of this assumption we obtain the correct expression for the integral energy principle.

Let us consider a spatially bounded, closed system, which, four-dimensionally, we may represent as a strip, outside of which the $T_{\mu\nu}$ vanish. Integrate equation (47c) over a space section. Since the integrals of $\frac{\partial T_{\mu 1}}{\partial x_1}$, $\frac{\partial T_{\mu 2}}{\partial x_2}$ and $\frac{\partial T_{\mu 3}}{\partial x_3}$ vanish because the $T_{\mu\nu}$ vanish at the limits of integration, we obtain

$$\frac{\partial}{\partial l}\left\{\int T_{\mu 4} dx_1 dx_2 dx_3\right\} = 0 \tag{49}$$

Inside the parentheses are the expressions for the momen-

* It has been attempted to remedy this lack of knowledge by considering the charged particles as proper singularities. But in my opinion this means giving up a real understanding of the structure of matter. It seems to me much better to admit our present inability rather than to be satisfied by a solution that is only apparent.

tum of the whole system, multiplied by i, together with the negative energy of the system, so that (49) expresses the conservation principles in their integral form. That this gives the right conception of energy and the conservation principles will be seen from the following considerations.

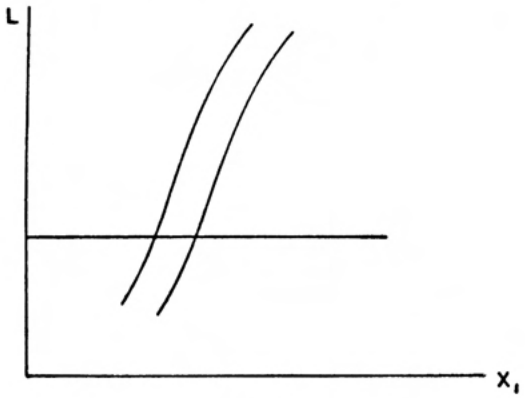


FIG. 3.

PHENOMENOLOGICAL REPRESENTATION OF THE ENERGY TENSOR OF MATTER

Hydrodynamical Equations. We know that matter is built up of electrically charged particles, but we do not know the laws which govern the constitution of these particles. In treating mechanical problems, we are therefore obliged to make use of an inexact description of matter, which corresponds to that of classical mechanics. The density σ, of a material substance and the hydrodynamical pressures are the fundamental concepts upon which such a description is based.

Let σ_0 be the density of matter at a place, estimated with reference to a system of co-ordinates moving with the matter. Then σ_0, the density at rest, is an invariant.

If we think of the matter in arbitrary motion and neglect the pressures (particles of dust *in vacuo*, neglecting the size of the particles and the temperature), then the energy tensor will depend only upon the velocity components, u_ν and σ_0. We secure the tensor character of $T_{\mu\nu}$ by putting

$$T_{\mu\nu} = \sigma_0 u_\mu u_\nu \tag{50}$$

in which the u_μ, in the three-dimensional representation, are given by (41). In fact, it follows from (50) that for $q = 0$, $T_{44} = -\sigma_0$ (equal to the negative energy per unit volume), as it should, according to the principle of the equivalence of mass and energy, and according to the physical interpretation of the energy tensor given above. If an external force (four-dimensional vector, K_μ) acts upon the matter, by the principles of momentum and energy the equation

$$K_\mu = \frac{\partial T_{\mu\nu}}{\partial x_\nu}$$

must hold. We shall now show that this equation leads to the same law of motion of a material particle as that already obtained. Let us imagine the matter to be of infinitely small extent in space, that is, a four-dimensional thread; then by integration over the whole thread with respect to the space co-ordinates x_1, x_2, x_3, we obtain

$$\int K_1 dx_1 dx_2 dx_3 = \int \frac{\partial T_{14}}{\partial x_4} dx_1 dx_2 dx_3 = -i \frac{d}{dl} \left\{ \int \sigma_0 \frac{dx_1}{d\tau} \frac{dx_4}{d\tau} dx_1 dx_2 dx_3 \right\}.$$

Now $\int dx_1 dx_2 dx_3 dx_4$ is an invariant, as is, therefore, also $\int \sigma_0 dx_1 dx_2 dx_3 dx_4$. We shall calculate this integral, first with respect to the inertial system which we have chosen,

and second, with respect to a system relatively to which the matter has the velocity zero. The integration is to be extended over a filament of the thread for which σ_0 may be regarded as constant over the whole section. If the space volumes of the filament referred to the two systems are dV and dV_0 respectively, then we have

$$\int \sigma_0 dV dl = \int \sigma_0 dV_0 d\tau$$

and therefore also

$$\int \sigma_0 dV = \int \sigma_0 dV_0 \frac{d\tau}{dl} = \int dm\, i \frac{d\tau}{dx_4}.$$

If we substitute the right-hand side for the left-hand side in the former integral, and put $\frac{dx_1}{d\tau}$ outside the sign of integration, we obtain,

$$K_x = \frac{d}{dl}\left(m \frac{dx_1}{d\tau} \right) = \frac{d}{dl}\left(\frac{m q_x}{\sqrt{1 - q^2}} \right).$$

We see, therefore, that the generalized conception of the energy tensor is in agreement with our former result.

The Eulerian Equations for Perfect Fluids. In order to get nearer to the behaviour of real matter we must add to the energy tensor a term which corresponds to the pressures. The simplest case is that of a perfect fluid in which the pressure is determined by a scalar p. Since the tangential stresses p_{xy}, etc., vanish in this case, the contribution to the energy tensor must be of the form $p\delta_{\mu\nu}$. We must therefore put

$$T_{\mu\nu} = \sigma u_\mu u_\nu + p\delta_{\mu\nu} \tag{51}$$

At rest, the density of the matter, or the energy per unit volume, is in this case, not σ but $\sigma - p$. For

$$-T_{44} = -\sigma \frac{dx_4}{d\tau}\frac{dx_4}{d\tau} - p\delta_{44} = \sigma - p.$$

In the absence of any force, we have

$$\frac{\partial T_{\mu\nu}}{\partial x_\nu} = \sigma u_\nu \frac{\partial u_\mu}{\partial x_\nu} + u_\mu \frac{\partial(\sigma u_\nu)}{\partial x_\nu} + \frac{\partial p}{\partial x_\mu} = 0.$$

If we multiply this equation by $u_\mu \left(= \frac{dx_\mu}{d\tau} \right)$ and sum for the μ's we obtain, using (40).

$$-\frac{\partial(\sigma u_\nu)}{\partial x_\nu} + \frac{dp}{d\tau} = 0 \tag{52}$$

where we have put $\frac{\partial p}{\partial x_\mu}\frac{dx_\mu}{d\tau} = \frac{dp}{d\tau}$. This is the equation of continuity, which differs from that of classical mechanics by the term $\frac{dp}{d\tau}$, which, practically, is vanishingly small. Observing (52), the conservation principles take the form

$$\sigma \frac{du_\mu}{d\tau} + u_\mu \frac{dp}{d\tau} + \frac{\partial p}{\partial x_\mu} = 0. \tag{53}$$

The equations for the first three indices evidently correspond to the Eulerian equations. That the equations (52) and (53) correspond, to a first approximation, to the hydrodynamical equations of classical mechanics, is a further confirmation of the generalized energy principle. The density of matter (or of energy) has tensor character (specifically, it constitutes a symmetrical tensor).

THE GENERAL THEORY OF RELATIVITY

ALL of the previous considerations have been based upon the assumption that all inertial systems are equivalent for the description of physical phenomena, but that they are preferred, for the formulation of the laws of nature, to spaces of reference in a different state of motion. We can think of no cause for this preference for definite states of motion to all others, according to our previous considerations, either in the perceptible bodies or in the concept of motion; on the contrary, it must be regarded as an independent property of the space-time continuum. The principle of inertia, in particular, seems to compel us to ascribe physically objective properties to the space-time continuum. Just as it was consistent from the Newtonian standpoint to make both the statements, *tempus est absolutum*, *spatium est absolutum*, so from the standpoint of the special theory of relativity we must say, *continuum spatii et temporis est absolutum*. In this latter statement *absolutum* means not only "physically real," but also "independent in its physical properties, having a physical effect, but not itself influenced by physical conditions."

As long as the principle of inertia is regarded as the keystone of physics, this standpoint is certainly the only one which is justified. But there are two serious criticisms of the ordinary conception. In the first place, it is contrary

to the mode of thinking in science to conceive of a thing (the space-time continuum) which acts itself, but which cannot be acted upon. This is the reason why E. Mach was led to make the attempt to eliminate space as an active cause in the system of mechanics. According to him, a material particle does not move in unaccelerated motion relatively to space, but relatively to the centre of all the other masses in the universe; in this way the series of causes of mechanical phenomena was closed, in contrast to the mechanics of Newton and Galileo. In order to develop this idea within the limits of the modern theory of action through a medium, the properties of the space-time continuum which determine inertia must be regarded as field properties of space, analogous to the electromagnetic field. The concepts of classical mechanics afford no way of expressing this. For this reason Mach's attempt at a solution failed for the time being. We shall come back to this point of view later. In the second place, classical mechanics exhibits a deficiency which directly calls for an extension of the principle of relativity to spaces of reference which are not in uniform motion relatively to each other. The ratio of the masses of two bodies is defined in mechanics in two ways which differ from each other fundamentally; in the first place, as the reciprocal ratio of the accelerations which the same motive force imparts to them (inert mass), and in the second place, as the ratio of the forces which act upon them in the same gravitational field (gravitational mass). The equality of these two masses, so differently defined, is a fact which is confirmed by experiments of very high accuracy (experiments of Eötvös), and classical mechanics offers no explanation for this equality. It is, however, clear that science is fully justified in assigning such a numerical equality only

after this numerical equality is reduced to an equality of the real nature of the two concepts.

That this object may actually be attained by an extension of the principle of relativity, follows from the following consideration. A little reflection will show that the law of the equality of the inert and the gravitational mass is equivalent to the assertion that the acceleration imparted to a body by a gravitational field is independent of the nature of the body. For Newton's equation of motion in a gravitational field, written out in full, is

(Inert mass) . (Acceleration) = (Intensity of the gravitational field) . (Gravitational mass).

It is only when there is numerical equality between the inert and gravitational mass that the acceleration is independent of the nature of the body. Let now K be an inertial system. Masses which are sufficiently far from each other and from other bodies are then, with respect to K, free from acceleration. We shall also refer these masses to a system of co-ordinates K', uniformly accelerated with respect to K. Relatively to K' all the masses have equal and parallel accelerations; with respect to K' they behave just as if a gravitational field were present and K' were unaccelerated. Overlooking for the present the question as to the "cause" of such a gravitational field, which will occupy us later, there is nothing to prevent our conceiving this gravitational field as real, that is, the conception that K' is "at rest" and a gravitational field is present we may consider as equivalent to the conception that only K is an "allowable" system of co-ordinates and no gravitational field is present. The assumption of the complete physical equivalence of the systems of coordinates, K and K', we call the "principle of equivalence;"

this principle is evidently intimately connected with the law of the equality between the inert and the gravitational mass, and signifies an extension of the principle of relativity to co-ordinate systems which are in non-uniform motion relatively to each other. In fact, through this conception we arrive at the unity of the nature of inertia and gravitation. For according to our way of looking at it, the same masses may appear to be either under the action of inertia alone (with respect to K) or under the combined action of inertia and gravitation (with respect to K'). The possibility of explaining the numerical equality of inertia and gravitation by the unity of their nature gives to the general theory of relativity, according to my conviction, such a superiority over the conceptions of classical mechanics, that all the difficulties encountered must be considered as small in comparison with this progress.

What justifies us in dispensing with the preference for inertial systems over all other co-ordinate systems, a preference that seems so securely established by experience? The weakness of the principle of inertia lies in this, that it involves an argument in a circle: a mass moves without acceleration if it is sufficiently far from other bodies; we know that it is sufficiently far from other bodies only by the fact that it moves without acceleration. Are there at all any inertial systems for very extended portions of the space-time continuum, or, indeed, for the whole universe? We may look upon the principle of inertia as established, to a high degree of approximation, for the space of our planetary system, provided that we neglect the perturbations due to the sun and planets. Stated more exactly, there are finite regions, where, with respect to a suitably chosen space of reference, material particles move freely

THE GENERAL THEORY

without acceleration, and in which the laws of the special theory of relativity, which have been developed above, hold with remarkable accuracy. Such regions we shall call "Galilean regions." We shall proceed from the consideration of such regions as a special case of known properties.

The principle of equivalence demands that in dealing with Galilean regions we may equally well make use of non-inertial systems, that is, such co-ordinate systems as, relatively to inertial systems, are not free from acceleration and rotation. If, further, we are going to do away completely with the vexing question as to the objective reason for the preference of certain systems of co-ordinates, then we must allow the use of arbitrarily moving systems of co-ordinates. As soon as we make this attempt seriously we come into conflict with that physical interpretation of space and time to which we were led by the special theory of relativity. For let K' be a system of co-ordinates whose z'-axis coincides with the z-axis of K, and which rotates about the latter axis with constant angular velocity. Are the configurations of rigid bodies, at rest relatively to K', in accordance with the laws of Euclidean geometry? Since K' is not an inertial system, we do not know directly the laws of configuration of rigid bodies with respect to K', nor the laws of nature, in general. But we do know these laws with respect to the inertial system K, and we can therefore infer their form with respect to K'. Imagine a circle drawn about the origin in the $x'y'$ plane of K', and a diameter of this circle. Imagine, further, that we have given a large number of rigid rods, all equal to each other. We suppose these laid in series along the periphery and the diameter of the circle, at rest relatively to K'. If U is the number of these rods along the periphery, D the number

along the diameter, then, if K' does not rotate relatively to K, we shall have

$$\frac{U}{D} = \pi.$$

But if K' rotates we get a different result. Suppose that at a definite time t, of K we determine the ends of all the rods. With respect to K all the rods upon the periphery experience the Lorentz contraction, but the rods upon the diameter do not experience this contraction (along their lengths!).* It therefore follows that

$$\frac{U}{D} > \pi.$$

It therefore follows that the laws of configuration of rigid bodies with respect to K' do not agree with the laws of configuration of rigid bodies that are in accordance with Euclidean geometry. If, further, we place two similar clocks (rotating with K'), one upon the periphery, and the other at the centre of the circle, then, judged from K, the clock on the periphery will go slower than the clock at the centre. The same thing must take place, judged from K', if we do not define time with respect to K' in a wholly unnatural way, (that is, in such a way that the laws with respect to K' depend explicitly upon the time). Space and time, therefore, cannot be defined with respect to K' as they were in the special theory of relativity with respect to inertial systems. But, according to the principle of equivalence, K' may also be considered as a system at rest, with respect to which there is a gravitational field (field of centrifugal force, and

* These considerations assume that the behavior of rods and clocks depends only upon velocities, and not upon accelerations, or, at least, that the influence of acceleration does not counteract that of velocity.

force of Coriolis). We therefore arrive at the result: the gravitational field influences and even determines the metrical laws of the space-time continuum. If the laws of configuration of ideal rigid bodies are to be expressed geometrically, then in the presence of a gravitational field the geometry is not Euclidean.

The case that we have been considering is analogous to that which is presented in the two-dimensional treatment of surfaces. It is impossible in the latter case also, to introduce co-ordinates on a surface (e.g. the surface of an ellipsoid) which have a simple metrical significance, while on a plane the Cartesian co-ordinates, x_1, x_2, signify directly lengths measured by a unit measuring rod. Gauss overcame this difficulty, in his theory of surfaces, by introducing curvilinear co-ordinates which, apart from satisfying conditions of continuity, were wholly arbitrary, and only afterwards these co-ordinates were related to the metrical properties of the surface. In an analogous way we shall introduce in the general theory of relativity arbitrary co-ordinates, x_1, x_2, x_3, x_4, which shall number uniquely the space-time points, so that neighbouring events are associated with neighbouring values of the co-ordinates; otherwise, the choice of co-ordinates is arbitrary. We shall be true to the principle of relativity in its broadest sense if we give such a form to the laws that they are valid in every such four-dimensional system of co-ordinates, that is, if the equations expressing the laws are co-variant with respect to arbitrary transformations.

The most important point of contact between Gauss's theory of surfaces and the general theory of relativity lies in the metrical properties upon which the concepts of both theories, in the main, are based. In the case of the theory of surfaces, Gauss's argument is as follows.

THE GENERAL THEORY

Plane geometry may be based upon the concept of the distance ds, between two infinitely near points. The concept of this distance is physically significant because the distance can be measured directly by means of a rigid measuring rod. By a suitable choice of Cartesian co-ordinates this distance may be expressed by the formula $ds^2 = dx_1^2 + dx_2^2$. We may base upon this quantity the concepts of the straight line as the geodesic ($\delta\int ds = 0$), the interval, the circle, and the angle, upon which the Euclidean plane geometry is built. A geometry may be developed upon another continuously curved surface, if we observe that an infinitesimally small portion of the surface may be regarded as plane, to within relatively infinitesimal quantities. There are Cartesian co-ordinates, X_1, X_2, upon such a small portion of the surface, and the distance between two points, measured by a measuring rod, is given by

$$ds^2 = dX_1^2 + dX_2^2.$$

If we introduce arbitrary curvilinear co-ordinates, x_1, x_2, on the surface, then dX_1, dX_2, may be expressed linearly in terms of dx_1, dx_2. Then everywhere upon the surface we have

$$ds^2 + g_{11}dx_1^2 + 2g_{12}dx_1dx_2 + g_{22}dx_2^2$$

where g_{11}, g_{12}, g_{22} are determined by the nature of the surface and the choice of co-ordinates; if these quantities are known, then it is also known how networks of rigid rods may be laid upon the surface. In other words, the geometry of surfaces may be based upon this expression for ds^2 exactly as plane geometry is based upon the corresponding expression.

There are analogous relations in the four-dimensional space-time continuum of physics. In the immediate

neighbourhood of an observer, falling freely in a gravitational field, there exists no gravitational field. We can therefore always regard an infinitesimally small region of the space-time continuum as Galilean. For such an infinitely small region there will be an inertial system (with the space co-ordinates, X_1, X_2, X_3, and the time co-ordinate X_4) relatively to which we are to regard the laws of the special theory of relativity as valid. The quantity which is directly measurable by our unit measuring rods and clocks,

$$dX_1^2 + dX_2^2 + dX_3^2 - dX_4^2$$

or its negative,

$$ds^2 = -dX_1^2 - dX_2^2 - dX_3^2 + dX_4^2 \tag{54}$$

is therefore a uniquely determinate invariant for two neighbouring events (points in the four-dimensional continuum), provided that we use measuring rods that are equal to each other when brought together and superimposed, and clocks whose rates are the same when they are brought together. In this the physical assumption is essential that the relative lengths of two measuring rods and the relative rates of two clocks are independent, in principle, of their previous history. But this assumption is certainly warranted by experience; if it did not hold there could be no sharp spectral lines, since the single atoms of the same element certainly do not have the same history, and since—on the assumption of relative variability of the single atoms depending on previous history—it would be absurd to suppose that the masses or proper frequencies of these atoms ever had been equal to one another.

Space-time regions of finite extent are, in general, not Galilean, so that a gravitational field cannot be done away

with by any choice of co-ordinates in a finite region. There is, therefore, no choice of co-ordinates for which the metrical relations of the special theory of relativity hold in a finite region. But the invariant ds always exists for two neighbouring points (events) of the continuum. This invariant ds may be expressed in arbitrary co-ordinates. If one observes that the local dX_ν may be expressed linearly in terms of the co-ordinate differentials dx_ν, ds^2 may be expressed in the form

$$ds^2 = g_{\mu\nu}dx_\mu dx_\nu. \tag{55}$$

The functions $g_{\mu\nu}$ describe, with respect to the arbitrarily chosen system of co-ordinates, the metrical relations of the space-time continuum and also the gravitational field. As in the special theory of relativity, we have to discriminate between time-like and space-like line elements in the four-dimensional continuum; owing to the change of sign introduced, time-like line elements have a real, space-like line elements an imaginary ds. The time-like ds can be measured directly by a suitably chosen clock.

According to what has been said, it is evident that the formulation of the general theory of relativity requires a generalization of the theory of invariants and the theory of tensors; the question is raised as to the form of the equations which are co-variant with respect to arbitrary point transformations. The generalized calculus of tensors was developed by mathematicians long before the theory of relativity. Riemann first extended Gauss's train of thought to continua of any number of dimensions; with prophetic vision he saw the physical meaning of this generalization of Euclid's geometry. Then followed the development of the theory in the form of the calculus of tensors, particularly by Ricci and Levi-Civita. This is

the place for a brief presentation of the most important mathematical concepts and operations of this calculus of tensors.

We designate four quantities, which are defined as functions of the x_ν with respect to every system of coordinates, as components, A^ν, of a contra-variant vector, if they transform in a change of co-ordinates as the co-ordinate differentials dx_ν. We therefore have

$$A^{\mu'} = \frac{\partial x'_\mu}{\partial x_\nu} A^\nu. \tag{56}$$

Besides these contra-variant vectors, there are also co-variant vectors. If B_ν are the components of a co-variant vector, these vectors are transformed according to the rule

$$B'_\mu = \frac{\partial x_\nu}{\partial x'_\mu} B_\nu. \tag{57}$$

The definition of a co-variant vector is chosen in such a way that a co-variant vector and a contra-variant vector together form a scalar according to the scheme,

$$\phi = B_\nu A^\nu \text{ (summed over the } \nu).$$

For we have

$$B'_\mu A^{\mu'} = \frac{\partial x_\alpha}{\partial x'_\mu} \frac{\partial x'_\mu}{\partial x_\beta} B_\alpha A^\beta = B_\alpha A^\alpha.$$

In particular, the derivatives $\frac{\partial \phi}{\partial x_\alpha}$ of a scalar ϕ, are components of a co-variant vector, which, with the co-ordinate differentials, form the scalar $\frac{\partial \phi}{\partial x_\alpha} dx_\alpha$; we see from this example how natural is the definition of the co-variant vectors.

There are here, also, tensors of any rank, which may have co-variant or contra-variant character with respect to

each index; as with vectors, the character is designated by the position of the index. For example, A_{μ}^{ν} denotes a tensor of the second rank, which is co-variant with respect to the index μ, and contra-variant with respect to the index ν. The tensor character indicates that the equation of transformation is

$$A_{\mu}^{\nu'} = \frac{\partial x_{\alpha}}{\partial x'_{\mu}} \frac{\partial x'_{\nu}}{\partial x_{\beta}} A_{\alpha}^{\beta}. \tag{58}$$

Tensors may be formed by the addition and subtraction of tensors of equal rank and like character, as in the theory of invariants of orthogonal linear substitutions, for example,

$$A_{\mu}^{\nu} + B_{\mu}^{\nu} = C_{\mu}^{\nu}. \tag{59}$$

The proof of the tensor character of C_{μ}^{ν} depends upon (58).

Tensors may be formed by multiplication, keeping the character of the indices, just as in the theory of invariants of linear orthogonal transformations, for example,

$$A_{\mu}^{\nu} B_{\sigma\tau} = C_{\mu\sigma\tau}^{\nu}. \tag{60}$$

The proof follows directly from the rule of transformation.

Tensors may be formed by contraction with respect to two indices of different character, for example,

$$A_{\mu\sigma\tau}^{\mu} = B_{\sigma\tau}. \tag{61}$$

The tensor character of $A_{\mu\sigma\tau}^{\mu}$ determines the tensor character of $B_{\sigma\tau}$. Proof—

$$A_{\mu\sigma\tau}^{\mu'} = \frac{\partial x_{\alpha}}{\partial x'_{\mu}} \frac{\partial x'_{\mu}}{\partial x_{\beta}} \frac{\partial x_{\delta}}{\partial x'_{\sigma}} \frac{\partial x_{\iota}}{\partial x'_{\tau}} A_{\alpha\delta\iota}^{\beta} = \frac{\partial x_{\delta}}{\partial x'_{\sigma}} \frac{\partial x_{\iota}}{\partial x'_{\tau}} A_{\alpha\delta\iota}^{\alpha}.$$

The properties of symmetry and skew-symmetry of a tensor with respect to two indices of like character have the same significance as in the theory of special relativity.

With this, everything essential has been said with regard to the algebraic properties of tensors.

THE GENERAL THEORY

The Fundamental Tensor. It follows from the invariance of ds^2 for an arbitrary choice of the dx_ν, in connexion with the condition of symmetry consistent with (55), that the $g_{\mu\nu}$ are components of a symmetrical co-variant tensor (Fundamental Tensor). Let us form the determinant, g, of the $g_{\mu\nu}$, and also the cofactors, divided by g, corresponding to the various $g_{\mu\nu}$. These cofactors, divided by g, will be denoted by $g^{\mu\nu}$, and their co-variant character is not yet known. Then we have

$$g_{\mu\alpha}g^{\mu\beta} = \delta_\alpha^\beta = \begin{matrix} 1 \text{ if } \alpha = \beta \\ 0 \text{ if } \alpha \neq \beta \end{matrix} \tag{62}$$

If we form the infinitely small quantities (co-variant vectors)

$$d\xi_\mu = g_{\mu\alpha}dx_\alpha \tag{63}$$

multiply by $g^{\mu\beta}$ and sum over the μ, we obtain, by the use of (62),

$$dx_\beta = g^{\beta\mu}d\xi_\mu. \tag{64}$$

Since the ratios of the $d\xi_\mu$ are arbitrary, and the dx_β as well as the $d\xi_\mu$ are components of vectors, it follows that the $g^{\mu\nu}$ are the components of a contra-variant tensor* (contra-variant fundamental tensor). The tensor character of δ_α^β (mixed fundamental tensor) accordingly follows,

* If we multiply (64) by $\frac{\partial x'_\alpha}{\partial x_\beta}$, sum over the β, and replace the $d\xi\mu$ by a transformation to the accented system, we obtain

$$dx'_\alpha = \frac{\partial x'_\sigma}{\partial x_\mu}\frac{\partial x'_\alpha}{\partial x_\beta}g^{\mu\beta}d\xi'_\sigma.$$

The statement made above follows from this, since, by (64), we must also have $dx'_\alpha = g^{\sigma\alpha'}d\xi'_\sigma$, and both equations must hold for every choice of the $d\xi'_\sigma$.

by (62). By means of the fundamental tensor, instead of tensors with co-variant index character, we can introduce tensors with contra-variant index character, and conversely. For example,

$$A^\mu = g^{\mu\alpha} A_\alpha$$
$$A_\mu = g_{\mu\alpha} A^\alpha$$
$$T^\sigma_\mu = g^{\sigma\nu} T_{\mu\nu}.$$

Volume Invariants. The volume element

$$\int dx_1 dx_2 dx_3 dx_4 = dx$$

is not an invariant. For by Jacobi's theorem,

$$dx' = \left| \frac{dx'_\mu}{dx_\nu} \right| dx. \tag{65}$$

But we can complement dx so that it becomes an invariant. If we form the determinant of the quantities

$$g'_{\mu\nu} = \frac{\partial x_\alpha}{\partial x'_\mu} \frac{\partial x_\beta}{\partial x'_\nu} g_{\alpha\beta}$$

we obtain, by a double application of the theorem of multiplication of determinants,

$$g' = |g'_{\mu\nu}| = \left| \frac{\partial x_\nu}{\partial x'_\mu} \right|^2 \cdot |g_{\mu\nu}| = \left| \frac{\partial x'_\mu}{\partial x_\nu} \right|^{-2} g.$$

We therefore get the invariant,

$$\sqrt{g'}\, dx' = \sqrt{g}\, dx.$$

Formation of Tensors by Differentiation. Although the algebraic operations of tensor formation have proved to be as simple as in the special case of invariance with respect to linear orthogonal transformations, nevertheless in the general case, the invariant differential operations

are, unfortunately, considerably more complicated. The reason for this is as follows. If A^μ is a contra-variant vector, the coefficients of its transformation, $\frac{\partial x'_\mu}{\partial x_\nu}$, are independent of position only if the transformation is a linear one. Then the vector components, $A^\mu + \frac{\partial A^\mu}{\partial x_\alpha} dx_\alpha$, at a neighbouring point transform in the same way as the A^μ, from which follows the vector character of the vector differentials, and the tensor character of $\frac{\partial A^\mu}{\partial x_\alpha}$. But if the $\frac{\partial x'_\mu}{\partial x_\nu}$ are variable this is no longer true.

That there are, nevertheless, in the general case, invariant differential operations for tensors, is recognized most satisfactorily in the following way, introduced by Levi-Civita and Weyl. Let (A^μ) be a contra-variant vector whose components are given with respect to the co-ordinate system of the x_ν. Let P_1 and P_2 be two infinitesimally near points of the continuum. For the infinitesimal region surrounding the point P_1, there is, according to our way of considering the matter, a co-ordinate system of the X_ν (with imaginary X_4-co-ordinate) for which the continuum is Euclidean. Let $A^\mu_{(1)}$ be the co-ordinates of the vector at the point P_1. Imagine a vector drawn at the point P_2, using the local system of the X_ν, with the same co-ordinates (parallel vector through P_2), then this parallel vector is uniquely determined by the vector at P_1 and the displacement. We designate this operation, whose uniqueness will appear in the sequel, the parallel displacement of the vector A_μ from P_1 to the infinitesimally near point P_2. If we form the vector difference of the vector (A^μ) at the point P_2 and the vector obtained by parallel displacement from

P_1 to P_2, we get a vector which may be regarded as the differential of the vector (A^ν) for the given displacement (dx_ν).

This vector displacement can naturally also be considered with respect to the co-ordinate system of the x_ν. If A^ν are the co-ordinates of the vector at P_1, $A^\nu + \delta A^\nu$ the co-ordinates of the vector displaced to P_2 along the interval (dx_ν), then the δA^ν do not vanish in this case. We know of these quantities, which do not have a vector character, that they must depend linearly and homogeneously upon the dx_ν and the A^ν. We therefore put

$$\delta A^\nu = -\Gamma^\nu_{\alpha\beta} A^\alpha dx_\beta. \tag{67}$$

In addition, we can state that the $\Gamma^\nu_{\alpha\beta}$ must be symmetrical with respect to the indices α and β. For we can assume from a representation by the aid of a Euclidean system of local co-ordinates that the same parallelogram will be described by the displacement of an element $d^{(1)}x_\nu$ along a second element $d^{(2)}x_\nu$ as by a displacement of $d^{(2)}x_\nu$ along $d^{(1)}x_\nu$. We must therefore have

$$d^{(2)}x_\nu + (d^{(1)}x_\nu - \Gamma^\nu_{\alpha\beta} d^{(1)}x_\alpha d^{(2)}x_\beta) = d^{(1)}x_\nu + (d^{(2)}x_\nu - \Gamma^\nu_{\alpha\beta} d^{(2)}x_\alpha d^{(1)}x_\beta).$$

The statement made above follows from this, after interchanging the indices of summation, α and β, on the right-hand side.

Since the quantities $g_{\mu\nu}$ determine all the metrical properties of the continuum, they must also determine the $\Gamma^\nu_{\alpha\beta}$. If we consider the invariant of the vector A^ν, that is, the square of its magnitude,

$$g_{\mu\nu} A^\mu A^\nu$$

which is an invariant, this cannot change in a parallel displacement. We therefore have

$$0 = \delta(g_{\mu\nu}A^\mu A^\nu) = \frac{\partial g_{\mu\nu}}{\partial x_\alpha} A^\mu A^\nu dx_\alpha + g_{\mu\nu}A^\mu \delta A^\nu + g_{\mu\nu}A^\nu \delta A^\mu$$

or, by (67),

$$\left(\frac{\partial g_{\mu\nu}}{\partial x_\alpha} - g_{\mu\beta}\Gamma^\beta_{\nu\alpha} - g_{\nu\beta}\Gamma^\beta_{\mu\alpha}\right) A^\mu A^\nu dx_\alpha = 0.$$

Owing to the symmetry of the expression in the brackets with respect to the indices μ and ν, this equation can be valid for an arbitrary choice of the vectors (A^μ) and dx, only when the expression in the brackets vanishes for all combinations of the indices. By a cyclic interchange of the indices μ, ν, a, we obtain thus altogether three equations, from which we obtain, on taking into account the symmetrical property of the $\Gamma^\alpha_{\mu\nu}$,

$$(68) \qquad \left[\begin{smallmatrix}\mu\nu\\ \alpha\end{smallmatrix}\right] = g_{\alpha\beta}\Gamma^\beta_{\mu\nu}$$

in which, following Christoffel, the abbreviation has been used,

$$(69) \qquad \left[\begin{smallmatrix}\mu\nu\\ \alpha\end{smallmatrix}\right] = \tfrac{1}{2}\left(\frac{\partial g_{\mu\alpha}}{\partial x_\nu} + \frac{\partial g_{\nu\alpha}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right).$$

If we multiply (68) by $g^{\alpha\sigma}$ and sum over the α, we obtain

$$(70) \qquad \Gamma^\sigma_{\mu\nu} = \tfrac{1}{2}g^{\sigma\alpha}\left(\frac{\partial g_{\mu\alpha}}{\partial x_\nu} + \frac{\partial g_{\nu\alpha}}{\partial x_\mu} - \frac{\partial g_{\mu\nu}}{\partial x_\alpha}\right) = \left\{\begin{smallmatrix}\mu\nu\\ \sigma\end{smallmatrix}\right\}$$

in which $\left\{\begin{smallmatrix}\mu\nu\\ \sigma\end{smallmatrix}\right\}$ is the Christoffel symbol of the second kind. Thus the quantities Γ are deduced from the $g_{\mu\nu}$. Equations (67) and (70) are the foundation for the following discussion.

THE GENERAL THEORY

Co-variant Differentiation of Tensors. If $(A^\mu + \delta A^\mu)$ is the vector resulting from an infinitesimal parallel displacement from P_1 to P_2, and $(A^\mu + dA^\mu)$ the vector A^μ at the point P_2, then the difference of these two,

$$dA^\mu - \delta A^\mu = \left(\frac{\partial A^\mu}{\partial x_\sigma} + \Gamma^\mu_{\sigma\alpha} A^\alpha\right) dx_\sigma$$

is also a vector. Since this is the case for an arbitrary choice of the dx_σ, it follows that

$$A^\mu_{:\sigma} = \frac{\partial A^\mu}{\partial x_\sigma} + \Gamma^\mu_{\sigma\alpha} A^\alpha \tag{71}$$

is a tensor, which we designate as the co-variant derivative of the tensor of the first rank (vector). Contracting this tensor, we obtain the divergence of the contra-variant tensor A^μ. In this we must observe that according to (70),

$$\Gamma^\sigma_{\mu\sigma} = \tfrac{1}{2} g^{\sigma\alpha} \frac{\partial g_{\sigma\alpha}}{\partial x_\mu} = \frac{1}{\sqrt{g}} \frac{\partial \sqrt{g}}{\partial x_\mu} \tag{72}$$

If we put, further,

$$A^\mu \sqrt{g} = \mathfrak{A}^\mu \tag{73}$$

a quantity designated by Weyl as the contra-variant tensor density* of the first rank, it follows that,

$$\mathfrak{A} = \frac{\partial \mathfrak{A}^\mu}{\partial x_\mu} \tag{74}$$

is a scalar density.

We get the law of parallel displacement for the co-variant vector B_μ by stipulating that the parallel displacement

* This expression is justified, in that $A^\mu \sqrt{g}\, dx = \mathfrak{A}^\mu dx$ has a tensor character. Every tensor, when multiplied by $\sqrt{g}$, changes into a tensor density. We employ capital Gothic letters for tensor densities.

shall be effected in such a way that the scalar

$$\phi = A^\mu B_\mu$$

remains unchanged, and that therefore

$$A^\mu \delta B_\mu + B_\mu \delta A^\mu$$

vanishes for every value assigned to (A^μ). We therefore get

$$\delta B_\mu = \Gamma^\alpha_{\mu\sigma} A_\alpha dx_\sigma. \tag{75}$$

From this we arrive at the co-variant derivative of the co-variant vector by the same process as that which led to (71),

$$B_{\mu\,;\,\sigma} = \frac{\partial B_\mu}{\partial x_\sigma} - \Gamma^\alpha_{\mu\sigma} B_\alpha. \tag{76}$$

By interchanging the indices μ and σ, and subtracting, we get the skew-symmetrical tensor,

$$\phi_{\mu\sigma} = \frac{\partial B_\mu}{\partial x_\sigma} - \frac{\partial B_\sigma}{\partial x_\mu}. \tag{77}$$

For the co-variant differentiation of tensors of the second and higher ranks we may use the process by which (75) was deduced. Let, for example, $(A_{\sigma\tau})$ be a co-variant tensor of the second rank. Then $A_{\sigma\tau}E^\sigma F^\tau$ is a scalar, if E and F are vectors. This expression must not be changed by the δ-displacement; expressing this by a formula, we get, using (67), $\delta A_{\sigma\tau}$, whence we get the desired co-variant derivative,

$$A_{\sigma\tau\,;\,\rho} = \frac{\partial A_{\sigma\tau}}{\partial x_\rho} - \Gamma^\alpha_{\sigma\rho} A_{\alpha\tau} - \Gamma^\alpha_{\tau\rho} A_{\sigma\alpha}. \tag{78}$$

In order that the general law of co-variant differentiation of tensors may be clearly seen, we shall write down two

co-variant derivatives deduced in an analogous way:

$$A^{\tau}_{\sigma;\rho} = \frac{\partial A^{\tau}_{\sigma}}{\partial x_{\rho}} - \Gamma^{\alpha}_{\sigma\rho} A^{\tau}_{\alpha} + \Gamma^{\tau}_{\alpha\rho} A^{\alpha}_{\sigma}. \tag{79}$$

$$A^{\sigma\tau}_{;\rho} = \frac{\partial A^{\sigma\tau}}{\partial x_{\rho}} + \Gamma^{\sigma}_{\alpha\rho} A^{\alpha\tau} + \Gamma^{\tau}_{\alpha\rho} A^{\sigma\alpha} \tag{80}$$

The general law of formation now becomes evident. From these formulæ we shall deduce some others which are of interest for the physical applications of the theory.

In case $A_{\sigma\tau}$ is skew-symmetrical, we obtain the tensor

$$A_{\sigma\tau\rho} = \frac{\partial A_{\sigma\tau}}{\partial x_{\rho}} + \frac{\partial A_{\tau\rho}}{\partial x_{\sigma}} + \frac{\partial A_{\rho\sigma}}{\partial x_{\tau}} \tag{81}$$

which is skew-symmetrical in all pairs of indices, by cyclic interchanges and addition.

If, in (78), we replace $A_{\sigma\tau}$ by the fundamental tensor, $g_{\sigma\tau}$, then the right-hand side vanishes identically; an analogous statement holds for (80) with respect to $g^{\sigma\tau}$; that is, the co-variant derivatives of the fundamental tensor vanish. That this must be so we see directly in the local system of co-ordinates.

In case $A^{\sigma\tau}$ is skew-symmetrical, we obtain from (80), by contraction with respect to τ and ρ,

$$\mathfrak{A}^{\sigma} = \frac{\partial \mathfrak{A}^{\sigma\tau}}{\partial x_{\tau}}. \tag{82}$$

In the general case, from (79) and (80), by contraction with respect to τ and ρ, we obtain the equations,

$$\mathfrak{A}_{\sigma} = \frac{\partial \mathfrak{A}^{\alpha}_{\sigma}}{\partial x_{\alpha}} - \Gamma^{\alpha}_{\sigma\beta} \mathfrak{A}^{\beta}_{\alpha}. \tag{83}$$

$$\mathfrak{A}^{\sigma} = \frac{\partial \mathfrak{A}^{\sigma\alpha}}{\partial x_{\alpha}} + \Gamma^{\sigma}_{\alpha\beta} \mathfrak{A}^{\alpha\beta}. \tag{84}$$

THE GENERAL THEORY

The Riemann Tensor. If we have given a curve extending from the point P to the point G of the continuum, then a vector A^μ, given at P, may, by a parallel displacement, be moved along the curve to G. If the continuum is Euclidean (more generally, if by a suitable choice of co-ordinates the $g_{\mu\nu}$ are constants) then the vector obtained at G as a result of this displacement does not depend upon the choice of the curve joining P and G. But otherwise, the result depends upon the path of the displacement. In this case,

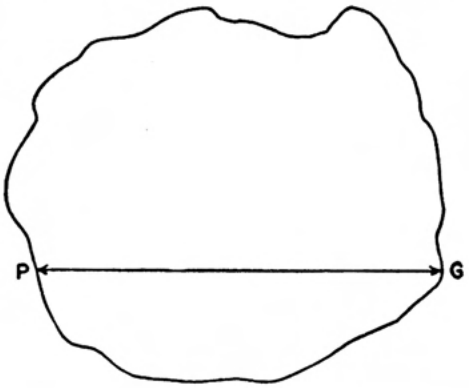


FIG. 4.

therefore, a vector suffers a change, ΔA^μ (in its direction, not its magnitude), when it is carried from a point P of a closed curve, along the curve, and back to P. We shall now calculate this vector change:

$$\Delta A^\mu = \oint \delta A^\mu.$$

As in Stokes' theorem for the line integral of a vector around a closed curve, this problem may be reduced to the integration around a closed curve with infinitely small linear dimensions; we shall limit ourselves to this case.

We have, first, by (67),

$$\Delta A^{\mu} = -\oint \Gamma^{\mu}_{\alpha\beta} A^{\alpha} dx_{\beta}.$$

In this, $\Gamma^{\mu}_{\alpha\beta}$ is the value of this quantity at the variable point G of the path of integration. If we put

$$\xi^{\mu} = (x_{\mu})_G - (x_{\mu})_P$$

and denote the value of $\Gamma^{\mu}_{\alpha\beta}$ at P by $\overline{\Gamma^{\mu}_{\alpha\beta}}$, then we have, with sufficient accuracy,

$$\Gamma^{\mu}_{\alpha\beta} = \overline{\Gamma^{\mu}_{\alpha\beta}} + \frac{\overline{\partial \Gamma^{\mu}_{\alpha\beta}}}{\partial x_{\tau}} \xi^{\tau}.$$

Let, further, A^{α} be the value obtained from $\overline{A^{\alpha}}$ by a parallel displacement along the curve from P to G. It may now easily be proved by means of (67) that $A^{\mu} - \overline{A^{\mu}}$ is infinitely small of the first order, while, for a curve of infinitely small dimensions of the first order, ΔA^{μ} is infinitely small of the second order. Therefore there is an error of only the second order if we put

$$A^{\alpha} = \overline{A^{\alpha}} - \overline{\Gamma^{\alpha}_{\sigma\tau}}\,\overline{A^{\sigma}}\,\overline{\xi^{\tau}}.$$

If we introduce these values of $\Gamma^{\mu}_{\alpha\beta}$ and A^{α} into the integral, we obtain, neglecting all quantities of a higher order than the second,

$$\Delta A^{\mu} = -\left(\frac{\partial \Gamma^{\mu}_{\sigma\beta}}{\partial x_{\alpha}} - \Gamma^{\mu}_{\rho\beta}\Gamma^{\rho}_{\sigma\alpha}\right) A^{\sigma} \oint \xi^{\alpha} d\xi^{\beta}. \tag{85}$$

The quantity removed from under the sign of integration refers to the point P. Subtracting $\frac{1}{2}d(\xi^{\alpha}\xi^{\beta})$ from the integrand, we obtain

$$\tfrac{1}{2} \oint (\xi^{\alpha} d\xi^{\beta} - \xi^{\beta} d\xi^{\alpha}).$$

This skew-symmetrical tensor of the second rank, $f^{\alpha\beta}$, characterizes the surface element bounded by the curve in magnitude and position. If the expression in the brackets in (85) were skew-symmetrical with respect to the indices α and β, we could conclude its tensor character from (85). We can accomplish this by interchanging the summation indices α and β in (85) and adding the resulting equation to (85). We obtain

$$2\Delta A^{\mu} = -R^{\mu}_{\sigma\alpha\beta}A^{\sigma}f^{\alpha\beta} \tag{86}$$

in which

$$R^{\mu}_{\sigma\alpha\beta} = -\frac{\partial\Gamma^{\mu}_{\sigma\alpha}}{\partial x_{\beta}} + \frac{\partial\Gamma^{\mu}_{\sigma\beta}}{\partial x_{\alpha}} + \Gamma^{\mu}_{\rho\alpha}\Gamma^{\rho}_{\sigma\beta} - \Gamma^{\mu}_{\rho\beta}\Gamma^{\rho}_{\sigma\alpha}. \tag{87}$$

The tensor character of $R^{\mu}_{\sigma\alpha\beta}$ follows from (86); this is the Riemann curvature tensor of the fourth rank, whose properties of symmetry we do not need to go into. Its vanishing is a sufficient condition (disregarding the reality of the chosen co-ordinates) that the continuum is Euclidean.

By contraction of the Riemann tensor with respect to the indices μ, β, we obtain the symmetrical tensor of the second rank,

$$R_{\mu\nu} = -\frac{\partial\Gamma^{\alpha}_{\mu\nu}}{\partial x_{\alpha}} + \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha} + \frac{\partial\Gamma^{\alpha}_{\mu\alpha}}{\partial x_{\nu}} - \Gamma^{\alpha}_{\mu\nu}\Gamma^{\beta}_{\alpha\beta}. \tag{88}$$

The last two terms vanish if the system of co-ordinates is so chosen that g = constant. From $R_{\mu\nu}$ we can form the scalar,

$$R = g^{\mu\nu}R_{\mu\nu}. \tag{89}$$

Straightest (Geodesic) Lines. A line may be constructed in such a way that its successive elements arise from each other by parallel displacements. This is the natural

generalization of the straight line of the Euclidean geometry. For such a line, we have

$$\delta\left(\frac{dx_\mu}{ds}\right) = -\Gamma^\mu_{\alpha\beta}\frac{dx_\alpha}{ds}dx_\beta.$$

The left-hand side is to be replaced by $\frac{d^2x_\mu}{ds^2}$,* so that we have

$$\frac{d^2x_\mu}{ds^2} + \Gamma^\mu_{\alpha\beta}\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds} = 0. \tag{90}$$

We get the same line if we find the line which gives a stationary value to the integral

$$\int ds \text{ or } \int \sqrt{g_{\mu\nu}dx_\mu dx_\nu}$$

between two points (geodesic line).

* The direction vector at a neighbouring point of the curve results, by a parallel displacement along the line element (dx_β), from the direction vector of each point considered.

THE GENERAL THEORY OF RELATIVITY (*Continued*)

WE are now in possession of the mathematical apparatus which is necessary to formulate the laws of the general theory of relativity. No attempt will be made in this presentation at systematic completeness, but single results and possibilities will be developed progressively from what is known and from the results obtained. Such a presentation is most suited to the present provisional state of our knowledge.

A material particle upon which no force acts moves, according to the principle of inertia, uniformly in a straight line. In the four-dimensional continuum of the special theory of relativity (with real time co-ordinate) this is a real straight line. The natural, that is, the simplest, generalization of the straight line which is meaningful in the system of concepts of the general (Riemannian) theory of invariants is that of the straightest, or geodesic, line. We shall accordingly have to assume, in the sense of the principle of equivalence, that the motion of a material particle, under the action only of inertia and gravitation, is described by the equation,

$$\frac{d^2x_\mu}{ds^2} + \Gamma^\mu_{\alpha\beta}\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds} = 0. \tag{90}$$

In fact, this equation reduces to that of a straight line if all the components, $\Gamma^\mu_{\alpha,\nu}$, of the gravitational field vanish.

THE GENERAL THEORY

How are these equations connected with Newton's equations of motion? According to the special theory of relativity, the $g_{\mu\nu}$ as well as the $g^{\mu\nu}$, have the values, with respect to an inertial system (with real time co-ordinate and suitable choice of the sign of ds^2),

$$(91) \qquad \begin{Bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{Bmatrix}$$

The equations of motion then become

$$\frac{d^2x_\mu}{ds^2} = 0.$$

We shall call this the "first approximation" to the $g_{\mu\nu}$-field. In considering approximations it is often useful, as in the special theory of relativity, to use an imaginary x_4-co-ordinate, as then the $g_{\mu\nu}$, to the first approximation, assume the values

$$(91a) \qquad \begin{Bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{Bmatrix}$$

These values may be collected in the relation

$$g_{\mu\nu} = -\delta_{\nu\mu}.$$

To the second approximation we must then put

$$(92) \qquad g_{\mu\nu} = -\delta_{\mu\nu} + \gamma_{\mu\nu}$$

where the $\gamma_{\mu\nu}$ are to be regarded as small of the first order.

Both terms of our equation of motion are then small of the first order. If we neglect terms which, relatively to these, are small of the first order, we have to put

$$ds^2 = -dx_\nu^2 = dl^2(1 - q^2) \tag{93}$$

$$\Gamma^{\mu}_{\alpha\beta} = -\delta_{\mu\sigma}\left[\begin{smallmatrix}\alpha\beta\\ \sigma\end{smallmatrix}\right] = -\left[\begin{smallmatrix}\alpha\beta\\ \mu\end{smallmatrix}\right] = \frac{1}{2}\left(\frac{\partial\gamma_{\alpha\beta}}{\partial x_\mu} - \frac{\partial\gamma_{\alpha\mu}}{\partial x_\beta} - \frac{\partial\gamma_{\beta\mu}}{\partial x_\alpha}\right). \tag{94}$$

We shall now introduce an approximation of a second kind. Let the velocity of the material particles be very small compared to that of light. Then ds will be the same as the time differential, dl. Further, $\frac{dx_1}{ds}$, $\frac{dx_2}{ds}$, $\frac{dx_3}{ds}$ will vanish compared to $\frac{dx_4}{ds}$. We shall assume, in addition, that the gravitational field varies so little with the time that the derivatives of the $\gamma_{\mu\nu}$ by x_4 may be neglected. Then the equation of motion (for $\mu = 1, 2, 3$) reduces to

$$\frac{d^2x_\mu}{dl^2} = \frac{\partial}{\partial x_\mu}\left(\frac{\gamma_{44}}{2}\right) \tag{90a}$$

This equation is identical with Newton's equation of motion for a material particle in a gravitational field, if we identify $\left(\frac{\gamma_{44}}{2}\right)$ with the potential of the gravitational field; whether or not this is allowable, naturally depends upon the field equations of gravitation, that is, it depends upon whether or not this quantity satisfies, to a first approximation, the same laws of the field as the gravitational potential in Newton's theory. A glance at (90) and (90a) shows that the $\Gamma^{\mu}_{\beta\alpha}$ actually do play the rôle of the intensity of the gravitational field. These quantities do not have a tensor character.

Equations (90) express the influence of inertia and gravitation upon the material particle. The unity of

inertia and gravitation is formally expressed by the fact that the whole left-hand side of (90) has the character of a tensor (with respect to any transformation of co-ordinates), but the two terms taken separately do not have tensor character. In analogy with Newton's equations, the first term would be regarded as the expression for inertia, and the second as the expression for the gravitational force.

We must next attempt to find the laws of the gravitational field. For this purpose, Poisson's equation,

$$\Delta\phi = 4\pi K\rho$$

of the Newtonian theory must serve as a model. This equation has its foundation in the idea that the gravitational field arises from the density ρ of ponderable matter. It must also be so in the general theory of relativity. But our investigations of the special theory of relativity have shown that in place of the scalar density of matter we have the tensor of energy per unit volume. In the latter is included not only the tensor of the energy of ponderable matter, but also that of the electromagnetic energy. We have seen, indeed, that in a more complete analysis the energy tensor can be regarded only as a provisional means of representing matter. In reality, matter consists of electrically charged particles, and is to be regarded itself as a part, in fact, the principal part, of the electromagnetic field. It is only the circumstance that we have no sufficient knowledge of the electromagnetic field of concentrated charges that compels us, provisionally, to leave undetermined in presenting the theory, the true form of this tensor. From this point of view it is at present appropriate to introduce a tensor $T_{\mu\nu}$ of the second rank of as yet unknown structure, which provisionally combines the energy density of the electromagnetic field and that of ponderable matter;

we shall denote this in the following as the "energy tensor of matter."

According to our previous results, the principles of momentum and energy are expressed by the statement that the divergence of this tensor vanishes (47c). In the general theory of relativity, we shall have to assume as valid the corresponding general co-variant equation. If $(T_{\mu\nu})$ denotes the co-variant energy tensor of matter, $\mathfrak{T}_\sigma^\nu$ the corresponding mixed tensor density, then, in accordance with (83), we must require that

$$0 = \frac{\partial \mathfrak{T}_\sigma^\alpha}{\partial x_\alpha} - \Gamma_{\sigma\beta}^\alpha \mathfrak{T}_\alpha^\beta \tag{95}$$

be satisfied. It must be remembered that besides the energy density of the matter there must also be given an energy density of the gravitational field, so that there can be no talk of principles of conservation of energy and momentum for matter alone. This is expressed mathematically by the presence of the second term in (95), which makes it impossible to conclude the existence of an integral equation of the form of (49). The gravitational field transfers energy and momentum to the "matter," in that it exerts forces upon it and gives it energy; this is expressed by the second term in (95).

If there is an analogue of Poisson's equation in the general theory of relativity, then this equation must be a tensor equation for the tensor $g_{\mu\nu}$ of the gravitational potential; the energy tensor of matter must appear on the right-hand side of this equation. On the left-hand side of the equation there must be a differential tensor in the $g_{\mu\nu}$. We have to find this differential tensor. It is completely determined by the following three conditions:

1. It may contain no differential coefficients of the $g_{\mu\nu}$ higher than the second.

2. It must be linear in these second differential coefficients.

3. Its divergence must vanish identically.

The first two of these conditions are naturally taken from Poisson's equation. Since it may be proved mathematically that all such differential tensors can be formed algebraically (i.e. without differentiation) from Riemann's tensor, our tensor must be of the form

$$R_{\mu\nu} + a g_{\mu\nu} R$$

in which $R_{\mu\nu}$ and R are defined by (88) and (89) respectively. Further, it may be proved that the third condition requires a to have the value $-\frac{1}{2}$. For the law of the gravitational field we therefore get the equation

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = -\kappa T_{\mu\nu}. \tag{96}$$

Equation (95) is a consequence of this equation. κ denotes a constant, which is connected with the Newtonian gravitation constant.

In the following I shall indicate the features of the theory which are interesting from the point of view of physics, using as little as possible of the rather involved mathematical method. It must first be shown that the divergence of the left-hand side actually vanishes. The energy principle for matter may be expressed, by (83),

$$0 = \frac{\partial \mathfrak{T}_\sigma^\alpha}{\partial x_\alpha} - \Gamma_{\sigma\beta}^\alpha \mathfrak{T}_\alpha^\beta \tag{97}$$

in which

$$\mathfrak{T}_\sigma^\alpha = T_{\sigma\tau} g^{\tau\alpha} \sqrt{-g}.$$

The analogous operation, applied to the left-hand side of (96), will lead to an identity.

THE GENERAL THEORY

In the region surrounding each world-point there are systems of co-ordinates for which, choosing the x_4-coordinate imaginary, at the given point,

$$g_{\mu\nu} = g^{\mu\nu} = -\delta_{\mu\nu} \begin{cases} = -1 \text{ if } \mu = \nu \\ = 0 \text{ if } \mu \neq \nu, \end{cases}$$

and for which the first derivatives of the $g_{\mu\nu}$ and the $g^{\mu\nu}$ vanish. We shall verify the vanishing of the divergence of the left-hand side at this point. At this point the components $\Gamma^{\alpha}_{\sigma\beta}$ vanish, so that we have to prove the vanishing only of

$$\frac{\partial}{\partial x_\sigma}[\sqrt{-g}\, g^{\nu\sigma}(R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R)].$$

Introducing (88) and (70) into this expression, we see that the only terms that remain are those in which third derivatives of the $g_{\mu\nu}$ enter. Since the $g_{\mu\nu}$ are to be replaced by $-\delta_{\mu\nu}$, we obtain, finally, only a few terms which may easily be seen to cancel each other. Since the quantity that we have formed has a tensor character, its vanishing is proved for every other system of co-ordinates also, and naturally for every other four-dimensional point. The energy principle of matter (97) is thus a mathematical consequence of the field equations (96).

In order to learn whether the equations (96) are consistent with experience, we must, above all else, find out whether they lead to the Newtonian theory as a first approximation. For this purpose we must introduce various approximations into these equations. We already know that Euclidean geometry and the law of the constancy of the velocity of light are valid, to a certain approximation, in regions of a great extent, as in the planetary sys-

tem. If, as in the special theory of relativity, we take the fourth co-ordinate imaginary, this means that we must put

$$g_{\mu\nu} = -\delta_{\mu\nu} + \gamma_{\mu\nu} \tag{98}$$

in which the $\gamma_{\mu\nu}$ are so small compared to 1 that we can neglect the higher powers of the $\gamma_{\mu\nu}$ and their derivatives. If we do this, we learn nothing about the structure of the gravitational field, or of metrical space of cosmical dimensions, but we do learn about the influence of neighbouring masses upon physical phenomena.

Before carrying through this approximation we shall transform (96). We multiply (96) by $g^{\mu\nu}$, summed over the μ and ν; observing the relation which follows from the definition of the $g^{\mu\nu}$,

$$g_{\mu\nu}g^{\mu\nu} = 4$$

we obtain the equation

$$R = \kappa g^{\mu\nu}T_{\mu\nu} = \kappa T.$$

If we put this value of R in (96) we obtain

$$R_{\mu\nu} = -\kappa(T_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}T) = -\kappa T^{*}_{\mu\nu}. \tag{96a}$$

When the approximation which has been mentioned is carried out, we obtain for the left-hand side,

$$-\tfrac{1}{2}\left(\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha^2} + \frac{\partial^2\gamma_{\alpha\alpha}}{\partial x_\mu \partial x_\nu} - \frac{\partial^2\gamma_{\mu\alpha}}{\partial x_\nu \partial x_\alpha} - \frac{\partial^2\gamma_{\nu\alpha}}{\partial x_\mu \partial x_\alpha}\right)$$

or

$$-\tfrac{1}{2}\frac{\partial^2\gamma_{\mu\nu}}{\partial x_\alpha^2} + \tfrac{1}{2}\frac{\partial}{\partial x_\nu}\left(\frac{\partial\gamma'_{\mu\alpha}}{\partial x_\alpha}\right) + \tfrac{1}{2}\frac{\partial}{\partial x_\mu}\left(\frac{\partial\gamma'_{\nu\alpha}}{\partial x_\alpha}\right)$$

in which has been put

$$\gamma'_{\mu\nu} = \gamma_{\mu\nu} - \tfrac{1}{2}\gamma_{\sigma\sigma}\delta_{\mu\nu}. \tag{99}$$

THE GENERAL THEORY

We must now note that equation (96) is valid for any system of co-ordinates. We have already specialized the system of co-ordinates in that we have chosen it so that within the region considered the $g_{\mu\nu}$ differ infinitely little from the constant values $-\delta_{\mu\nu}$. But this condition remains satisfied in any infinitesimal change of co-ordinates, so that there are still four conditions to which the $\gamma_{\mu\nu}$ may be subjected, provided these conditions do not conflict with the conditions for the order of magnitude of the $\gamma_{\mu\nu}$. We shall now assume that the system of co-ordinates is so chosen that the four relations—

$$0 = \frac{\partial \gamma'_{\mu\nu}}{\partial x_\nu} = \frac{\partial \gamma_{\mu\nu}}{\partial x_\nu} - \frac{1}{2}\frac{\partial \gamma_{\sigma\sigma}}{\partial x_\mu} \tag{100}$$

are satisfied. Then (96a) takes the form

$$\frac{\partial^2 \gamma_{\mu\nu}}{\partial x_\alpha{}^2} = 2\kappa T^*_{\mu\nu} \tag{96b}$$

These equations may be solved by the method, familiar in electrodynamics, of retarded potentials; we get, in an easily understood notation,

$$\gamma_{\mu\nu} = -\frac{\kappa}{2\pi}\int \frac{T^*_{\mu\nu}(x_0, y_0, z_0, t - r)}{r}\, dV_0. \tag{101}$$

In order to see in what sense this theory contains the Newtonian theory, we must consider in greater detail the energy tensor of matter. Considered phenomenologically, this energy tensor is composed of that of the electromagnetic field and of matter in the narrower sense. If we consider the different parts of this energy tensor with respect to their order of magnitude, it follows from the results of the special theory of relativity that the contribution of the electromagnetic field practically vanishes in comparison

to that of ponderable matter. In our system of units, the energy of one gram of matter is equal to 1, compared to which the energy of the electric fields may be ignored, and also the energy of deformation of matter, and even the chemical energy. We get an approximation that is fully sufficient for our purpose if we put

$$(102) \qquad \begin{cases} T^{\mu\nu} = \sigma \dfrac{dx_\mu}{ds} \dfrac{dx_\nu}{ds} \\ ds^2 = g_{\mu\nu} dx_\mu dx_\nu \end{cases}$$

In this, σ is the density at rest, that is, the density of the ponderable matter, in the ordinary sense, measured with the aid of a unit measuring rod, and referred to a Galilean system of co-ordinates moving with the matter.

We observe, further, that in the co-ordinates we have chosen, we shall make only a relatively small error if we replace the $g_{\mu\nu}$ by $-\delta_{\mu\nu}$, so that we put

$$(102a) \qquad ds^2 = -\sum dx_\mu^2.$$

The previous developments are valid however rapidly the masses which generate the field may move relatively to our chosen system of quasi-Galilean co-ordinates. But in astronomy we have to do with masses whose velocities, relatively to the co-ordinate system employed, are always small compared to the velocity of light, that is, small compared to 1, with our choice of the unit of time. We therefore get an approximation which is sufficient for nearly all practical purposes if in (101) we replace the retarded potential by the ordinary (non-retarded) potential, and if, for the masses which generate the field, we put

$$(103a) \qquad \frac{dx_1}{ds} = \frac{dx_2}{ds} = \frac{dx_3}{ds} = 0, \quad \frac{dx_4}{ds} = \frac{\sqrt{-1}\, dl}{dl} = \sqrt{-1}.$$

Then we get for $T^{\mu\nu}$ and $T_{\mu\nu}$ the values

$$(104) \qquad \begin{cases} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\sigma \end{cases}$$

For T we get the value σ, and, finally, for $T^{*}_{\mu\nu}$ the values,

$$(104a) \qquad \begin{cases} \frac{\sigma}{2} & 0 & 0 & 0 \\ 0 & \frac{\sigma}{2} & 0 & 0 \\ 0 & 0 & \frac{\sigma}{2} & 0 \\ 0 & 0 & 0 & -\frac{\sigma}{2} \end{cases}$$

We thus get, from (101),

$$(101a) \qquad \begin{cases} \gamma_{11} = \gamma_{22} = \gamma_{33} = -\frac{\kappa}{4\pi}\int \frac{\sigma dV_0}{r} \\ \gamma_{44} = +\frac{\kappa}{4\pi}\int \frac{\sigma dV_0}{r} \end{cases}$$

while at the other $\gamma_{\mu\nu}$ vanish. The last of these equations, in connexion with equation (90a), contains Newton's theory of gravitation. If we replace l by ct we get

$$(90b) \qquad \left\{ \frac{d^2x_\mu}{dt^2} = \frac{\kappa c^2}{8\pi}\frac{\partial}{\partial x_\mu}\int \frac{\sigma dV_0}{r}. \right.$$

We see that the Newtonian gravitation constant K, is connected with the constant κ that enters into our field equations by the relation

$$(105) \qquad K = \frac{\kappa c^2}{8\pi}.$$

From the known numerical value of K, it therefore follows that

$$(105a) \quad \kappa = \frac{8\pi K}{c^2} = \frac{8\pi \cdot 6{\cdot}67 \cdot 10^{-8}}{9 \cdot 10^{20}} = 1{\cdot}86 \cdot 10^{-27}.$$

From (101) we see that even in the first approximation the structure of the gravitational field differs fundamentally from that which is consistent with the Newtonian theory; this difference lies in the fact that the gravitational potential has the character of a tensor and not a scalar. This was not recognized in the past because only the component g_{44}, to a first approximation, enters the equations of motion of material particles.

In order now to be able to judge the behaviour of measuring rods and clocks from our results, we must observe the following. According to the principle of equivalence, the metrical relations of the Euclidean geometry are valid relatively to a Cartesian system of reference of infinitely small dimensions, and in a suitable state of motion (freely falling, and without rotation). We can make the same statement for local systems of co-ordinates which, relatively to these, have small accelerations, and therefore for such systems of co-ordinates as are at rest relatively to the one we have selected. For such a local system, we have, for two neighbouring point events,

$$ds^2 = -dX_1^2 - dX_2^2 - dX_3^2 + dT^2 = -dS^2 + dT^2$$

where dS is measured directly by a measuring rod and dT by a clock at rest relatively to the system: these are the naturally measured lengths and times. Since ds^2, on the other hand, is known in terms of the co-ordinates x_ν employed in finite regions, in the form

$$ds^2 = g_{\mu\nu} dx_\mu dx_\nu$$

we have the possibility of getting the relation between naturally measured lengths and times, on the one hand, and the corresponding differences of co-ordinates, on the other hand. As the division into space and time is in agreement with respect to the two systems of co-ordinates, so when we equate the two expressions for ds^2 we get two relations. If, by (101a), we put

$$ds^2 = -\left(1 + \frac{\kappa}{4\pi}\int \frac{\sigma dV_0}{r}\right)(dx_1^2 + dx_2^2 + dx_3^2) + \left(1 - \frac{\kappa}{4\pi}\int \frac{\sigma dV_0}{r}\right) dl^2$$

we obtain, to a sufficiently close approximation,

$$(106)\quad \begin{cases} \sqrt{dX_1^2 + dX_2^2 + dX_3^2} \\ \qquad = \left(1 + \frac{\kappa}{8\pi}\int \frac{\sigma dV_0}{r}\right)\sqrt{dx_1^2 + dx_2^2 + dx_3^2} \\ dT = \left(1 - \frac{\kappa}{8\pi}\int \frac{\sigma dV_0}{r}\right) dl. \end{cases}$$

The unit measuring rod has therefore the coordinate length,

$$1 - \frac{\kappa}{8\pi}\int \frac{\sigma dV_0}{r}$$

in respect to the system of co-ordinates we have selected. The particular system of co-ordinates we have selected insures that this length shall depend only upon the place, and not upon the direction. If we had chosen a different system of co-ordinates this would not be so. But however we may choose a system of co-ordinates, the laws of configuration of rigid rods do not agree with those of Euclidean geometry; in other words, we cannot choose any system of co-ordinates so that the co-ordinate differences, Δx_1,

Δx_2, Δx_3, corresponding to the ends of a unit measuring rod, oriented in any way, shall always satisfy the relation $\Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2 = 1$. In this sense space is not Euclidean, but "curved." It follows from the second of the relations above that the interval between two beats of the unit clock ($dT = 1$) corresponds to the "time"

$$1 + \frac{\kappa}{8\pi} \int \frac{\sigma dV_0}{r}$$

in the unit used in our system of co-ordinates. The rate of a clock is accordingly slower the greater is the mass of the ponderable matter in its neighbourhood. We therefore conclude that spectral lines which are produced on the sun's surface will be displaced towards the red, compared to the corresponding lines produced on the earth, by about $2 \cdot 10^{-6}$ of their wave-lengths. At first, this important consequence of the theory appeared to conflict with experiment; but results obtained during the past years seem to make the existence of this effect more and more probable, and it can hardly be doubted that this consequence of the theory will be confirmed within the next years.

Another important consequence of the theory, which can be tested experimentally, has to do with the path of rays of light. In the general theory of relativity also the velocity of light is everywhere the same, relatively to a local inertial system. This velocity is unity in our natural measure of time. The law of the propagation of light in general co-ordinates is therefore, according to the general theory of relativity, characterized, by the equation

$$ds^2 = 0.$$

To within the approximation which we are using, and in the system of co-ordinates which we have selected, the

velocity of light is characterized, according to (106), by the equation

$$\left(1 + \frac{\kappa}{4\pi}\int \frac{\sigma dV_0}{r}\right)(dx_1^2 + dx_2^2 + dx_3^2) = \left(1 - \frac{\kappa}{4\pi}\int \frac{\sigma dV_0}{r}\right) dl^2.$$

The velocity of light L, is therefore expressed in our co-ordinates by

$$\frac{\sqrt{dx_1^2 + dx_2^2 + dx_3^2}}{dl} = 1 - \frac{\kappa}{4\pi}\int \frac{\sigma dV_0}{r} \tag{107}$$

We can therefore draw the conclusion from this, that a ray of light passing near a large mass is deflected. If we imagine the sun, of mass M, concentrated at the origin of our system of co-ordinates, then a ray of light, travelling parallel to the x_3-axis, in the $x_1 - x_3$ plane, at a distance Δ from the origin, will be deflected, in all, by an amount

$$\alpha = \int_{-\infty}^{+\infty} \frac{1}{L}\frac{\partial L}{\partial x_1} dx_3$$

towards the sun. On performing the integration we get

$$\alpha = \frac{\kappa M}{2\pi\Delta}. \tag{108}$$

The existence of this deflection, which amounts to 1.7″ for Δ equal to the radius of the sun, was confirmed, with remarkable accuracy, by the English Solar Eclipse Expedition in 1919, and most careful preparations have been made to get more exact observational data at the solar eclipse in 1922. It should be noted that this result, also, of the theory is not influenced by our arbitrary choice of a system of co-ordinates.

THE GENERAL THEORY

This is the place to speak of the third consequence of the theory which can be tested by observation, namely, that which concerns the motion of the perihelion of the planet Mercury. The secular changes in the planetary orbits are known with such accuracy that the approximation we have been using is no longer sufficient for a comparison of theory and observation. It is necessary to go back to the general field equations (96). To solve this problem I made use of the method of successive approximations. Since then, however, the problem of the central symmetrical statical gravitational field has been completely solved by Schwarzschild and others; the derivation given by H. Weyl in his book, "Raum-Zeit-Materie," is particularly elegant. The calculation can be simplified somewhat if we do not go back directly to the equation (96), but base it upon a principle of variation that is equivalent to this equation. I shall indicate the procedure only in so far as is necessary for understanding the method.

In the case of a statical field, ds^2 must have the form

$$\begin{cases} ds^2 = -d\sigma^2 + f^2 dx_4^2 \\ d\sigma^2 = \sum_{1-3} \gamma_{\alpha\beta} dx_\alpha dx_\beta \end{cases} \tag{109}$$

where the summation on the right-hand side of the last equation is to be extended over the space variables only, The central symmetry of the field requires the $\gamma_{\mu\nu}$ to be of the form,

$$\gamma_{\alpha\beta} = \mu\delta_{\alpha\beta} + \lambda x_\alpha x_\beta \tag{110}$$

f^2, μ and λ are functions of $r = \sqrt{x_1^2 + x_2^2 + x_3^2}$ only. One of these three functions can be chosen arbitrarily, because our system of co-ordinates is, *a priori*, completely arbitrary; for by a substitution

$$x'_4 = x_4$$
$$x'_\alpha = F(r)x_\alpha$$

we can always insure that one of these three functions shall be an assigned function of r'. In place of (110) we can therefore put, without limiting the generality,

$$\gamma_{\alpha\beta} = \delta_{\alpha\beta} + \lambda x_\alpha x_\beta. \tag{110a}$$

In this way the $g_{\mu\nu}$ are expressed in terms of the two quantities λ and f. These are to be determined as functions of r, by introducing them into equation (96), after first calculating the $\Gamma^\sigma_{\mu\nu}$ from (109) and (110a). We have

$$\text{(110b)} \quad \begin{cases} \Gamma^\sigma_{\alpha\beta} = \frac{1}{2}\dfrac{x_\sigma}{r}\cdot\dfrac{\lambda' x_\alpha x_\beta + 2\lambda r\delta_{\alpha\beta}}{1+\lambda r^2} \text{ (for } \alpha, \beta, \sigma = 1, 2, 3) \\ \Gamma^4_{44} = \Gamma^\alpha_{4\beta} = \Gamma^4_{\alpha\beta} = 0 \text{ (for } \alpha, \beta = 1, 2, 3) \\ \Gamma^4_{4\alpha} = \frac{1}{2}f^{-2}\dfrac{\partial f^2}{\partial x_\alpha}, \qquad \Gamma^\alpha_{44} = -\frac{1}{2}g^{\alpha\beta}\dfrac{\partial f^2}{\partial x_\beta} \end{cases}$$

With the help of these results, the field equations furnish Schwarzschild's solution:

$$\text{(109a)} \quad ds^2 = \left(1 - \frac{A}{r}\right)dl^2 - \left[\frac{dr^2}{1-\dfrac{A}{r}} + r^2(\sin^2\theta d\phi^2 + d\theta^2)\right]$$

in which we have put

$$\text{(109b)} \quad \begin{cases} x_4 = l \\ x_1 = r\sin\theta\sin\phi \\ x_2 = r\sin\theta\cos\phi \\ x_3 = r\cos\theta \\ A = \dfrac{\kappa M}{4\pi} \end{cases}$$

THE GENERAL THEORY

M denotes the sun's mass, centrally symmetrically placed about the origin of co-ordinates; the solution (109a) is valid only outside of this mass, where all the $T_{\mu\nu}$ vanish. If the motion of the planet takes place in the $x_1 - x_2$ plane then we must replace (109a) by

$$(109c) \qquad ds^2 = \left(1 - \frac{A}{r}\right) dl^2 - \frac{dr^2}{1 - \frac{A}{r}} - r^2 d\phi^2.$$

The calculation of the planetary motion depends upon equation (90). From the first of equations (110b) and (90) we get, for the indices 1, 2, 3,

$$\frac{d}{ds}\left(x_\alpha \frac{dx_\beta}{ds} - x_\beta \frac{dx_\alpha}{ds}\right) = 0$$

or, if we integrate, and express the result in polar co-ordinates,

$$(111) \qquad r^2 \frac{d\phi}{ds} = \text{constant.}$$

From (90), for $\mu = 4$, we get

$$0 = \frac{d^2 l}{ds^2} + \frac{1}{f^2} \frac{\partial f^2}{\partial x_\alpha} \frac{dx_\alpha}{ds} \frac{dl}{ds} = \frac{d^2 l}{ds^2} + \frac{1}{f^2} \frac{df^2}{ds} \frac{dl}{ds}.$$

From this, after multiplication by f^2 and integration, we have

$$(112) \qquad f^2 \frac{dl}{ds} = \text{constant.}$$

In (109c), (111) and (112) we have three equations between the four variables s, r, l and ϕ, from which the motion of the planet may be calculated in the same way as in classical mechanics. The most important result we

get from this is a secular rotation of the elliptic orbit of the planet in the same sense as the revolution of the planet, amounting in radians per revolution to

$$\frac{24\pi^3 a^2}{(1-e^2)\,c^2 T^2} \tag{113}$$

where

a = the semi-major axis of the planetary orbit in centimetres.

e = the numerical eccentricity.

c = $3 \cdot 10^{+10}$, the velocity of the light *in vacuo*.

T = the period of revolution in seconds.

This expression furnishes the explanation of the motion of the perihelion of the planet Mercury, which has been known for a hundred years (since Leverrier), and for which theoretical astronomy has hitherto been unable satisfactorily to account.

There is no difficulty in expressing Maxwell's theory of the electromagnetic field in terms of the general theory of relativity; this is done by application of the tensor formation (81), (82) and (77). Let ϕ_μ be a tensor of the first rank, to be interpreted as an electromagnetic 4-potential; then as electromagnetic field tensor may be defined by the relations,

$$\phi_{\mu\nu} = \frac{\partial \phi_\mu}{\partial x_\mu} - \frac{\partial \phi_\nu}{\partial x_\mu}. \tag{114}$$

The second of Maxwell's systems of equations is then defined by the tensor equation, resulting from this,

$$\frac{\partial \phi_{\mu\nu}}{\partial x_\rho} + \frac{\partial \phi_{\nu\rho}}{\partial x_\mu} + \frac{\partial \phi_{\rho\mu}}{\partial x_\nu} = 0 \tag{114a}$$

and the first of Maxwell's systems of equations is defined by the tensor-density relation

THE GENERAL THEORY

$$\frac{\partial \mathfrak{F}^{\mu\nu}}{\partial x_\nu} = \mathfrak{T}^\mu \tag{115}$$

in which

$$\mathfrak{F}^{\mu\nu} = \sqrt{-g}\, g^{\mu\sigma} g^{\nu\tau} \phi_{\sigma\tau}$$

$$\mathfrak{T}^\mu = \sqrt{-g}\, \rho \frac{dx_\mu}{ds}.$$

If we introduce the energy tensor of the electromagnetic field into the right-hand side of (96), we obtain (115), for the special case $\mathfrak{T}^\mu = 0$, as a consequence of (96) by taking the divergence. This inclusion of the theory of electricity in the scheme of the general theory of relativity has been considered arbitrary and unsatisfactory by many theoreticians. Nor can we in this way understand the equilibrium of the electricity which constitutes the elementary electrically charged particles. A theory in which the gravitational field and the electromagnetic field do not enter as logically distinct structures would be much preferable. H. Weyl, and recently Th. Kaluza, have put forward ingenious ideas along this direction; but concerning them, I am convinced that they do not bring us nearer to the true solution of the fundamental problem. I shall not go into this further, but shall give a brief discussion of the so-called cosmological problem, for without this, the considerations regarding the general theory of relativity would, in a certain sense, remain unsatisfactory.

Our previous considerations, based upon the field equations (96), had for a foundation the conception that space on the whole is Galilean-Euclidean, and that this character is disturbed only by masses embedded in it. This conception was certainly justified as long as we were dealing with spaces of the order of magnitude of those that astronomy has mostly to do with. But whether portions of the

universe, however large they may be, are quasi-Euclidean, is a wholly different question. We can make this clear by using an example from the theory of surfaces which we have employed many times. If a certain portion of a surface is practically plane, it does not at all follow that the whole surface has the form of a plane; the surface might just as well be a sphere of sufficiently large radius. The question as to whether the universe as a whole is non-Euclidean was much discussed from the geometrical point of view before the development of the theory of relativity. But with the theory of relativity, this problem has entered upon a new stage, for according to this theory the geometrical properties of bodies are not independent, but depend upon the distribution of masses.

If the universe were quasi-Euclidean, then Mach was wholly wrong in his thought that inertia, as well as gravitation, depends upon a kind of mutual action between bodies. For in this case, for a suitably selected system of co-ordinates, the $g_{\mu\nu}$ would be constant at infinity, as they are in the special theory of relativity, while within finite regions the $g_{\mu\nu}$ would differ from these constant values by small amounts only, for a suitable choice co-ordinates, as a result of the influence of the masses in finite regions. The physical properties of space would not then be wholly independent, that is, uninfluenced by matter, but in the main they would be, and only in small measure conditioned by matter. Such a dualistic conception is even in itself not satisfactory; there are, however, some important physical arguments against it, which we shall consider.

The hypothesis that the universe is infinite and Euclidean at infinity, is, from the relativistic point of view, a complicated hypothesis. In the language of the general theory of relativity it demands that the Riemann tensor of

the fourth rank, R_{iklm}, shall vanish at infinity, which furnishes twenty independent conditions, while only ten curvature components, $R_{\mu\nu}$, enter into the laws of the gravitational field. It is certainly unsatisfactory to postulate such a far-reaching limitation without any physical basis for it.

But in the second place, the theory of relativity makes it appear probable that Mach was on the right road in his thought that inertia depends upon a mutual action of matter. For we shall show in the following that, according to our equations, inert masses do act upon each other in the sense of the relativity of inertia, even if only very feebly. What is to be expected along the line of Mach's thought?

1. The inertia of a body must increase when ponderable masses are piled up in its neighbourhood.
2. A body must experience an accelerating force when neighbouring masses are accelerated, and, in fact, the force must be in the same direction as that acceleration.
3. A rotating hollow body must generate inside of itself a "Coriolis field," which deflects moving bodies in the sense of the rotation, and a radial centrifugal field as well.

We shall now show that these three effects, which are to be expected in accordance with Mach's ideas, are actually present according to our theory, although their magnitude is so small that confirmation of them by laboratory experiments is not to be thought of. For this purpose we shall go back to the equations of motion of a material particle (90), and carry the approximations somewhat further than was done in equation (90a).

THE GENERAL THEORY

First, we consider γ_{44} as small of the first order. The square of the velocity of masses moving under the influence of the gravitational force is of the same order, according to the energy equation. It is therefore logical to regard the velocities of the material particles we are considering, as well as the velocities of the masses which generate the field, as small, of the order $\frac{1}{2}$. We shall now carry out the approximation in the equations that arise from the field equations (101) and the equations of motion (90) so far as to consider terms, in the second member of (90), that are linear in those velocities. Further, we shall not put ds and dl equal to each other, but, corresponding to the higher approximation, we shall put

$$ds = \sqrt{g_{44}}\, dl = \left(1 - \frac{\gamma_{44}}{2}\right) dl.$$

From (90) we obtain, at first,

$$(116)\quad \frac{d}{dl}\left[\left(1 + \frac{\gamma_{44}}{2}\right)\frac{dx_\mu}{dl}\right] = -\Gamma^{\mu}_{\alpha\beta}\frac{dx_\alpha}{dl}\frac{dx_\beta}{dl}\left(1 + \frac{\gamma_{44}}{2}\right).$$

From (101) we get, to the approximation sought for,

$$(117)\quad \begin{cases} -\gamma_{11} = -\gamma_{22} = -\gamma_{33} = \gamma_{44} = \dfrac{\kappa}{4\pi}\displaystyle\int \frac{\sigma dV_0}{r} \\ \gamma_{4\alpha} = -\dfrac{i\kappa}{2\pi}\displaystyle\int \sigma \frac{\dfrac{dx_\alpha}{ds}\, dV_0}{r} \\ \gamma_{\alpha\beta} = 0 \end{cases}$$

in which, in (117), α and β denote the space indices only.

On the right-hand side of (116) we can replace $1 + \frac{\gamma_{44}}{2}$ by 1 and $-\Gamma^{\alpha\beta}_{\mu}$ by $\left[{}^{\alpha\beta}_{\mu}\right]$. It is easy to see, in addition, that to this degree of approximation we must put

$$\begin{bmatrix} 44 \\ \mu \end{bmatrix} = -\frac{1}{2}\frac{\partial \gamma_{44}}{\partial x_\mu} + \frac{\partial \gamma_{4\mu}}{\partial x_4}$$

$$\begin{bmatrix} \alpha 4 \\ \mu \end{bmatrix} = \frac{1}{2}\left(\frac{\partial \gamma_{4\mu}}{\partial x_\alpha} - \frac{\partial \gamma_{4\alpha}}{\partial x_\mu}\right)$$

$$\begin{bmatrix} \alpha\beta \\ \mu \end{bmatrix} = 0$$

in which α, β and μ denote space indices. We therefore obtain from (116), in the usual vector notation,

$$(118)\quad \begin{cases} \dfrac{d}{dl}[(1+\bar{\sigma})\mathbf{v}] = \operatorname{grad}\bar{\sigma} + \dfrac{\partial \mathfrak{A}}{\partial l} + [\operatorname{curl}\mathfrak{A}, \mathbf{v}] \\ \bar{\sigma} = \dfrac{\kappa}{8\pi}\displaystyle\int \frac{\sigma dV_0}{r} \\ \mathfrak{A} = \dfrac{\kappa}{2\pi}\displaystyle\int \frac{\sigma \dfrac{dx_\alpha}{dt} dV_0}{r} \end{cases}$$

The equations of motion, (118), show now, in fact, that

1. The inert mass is proportional to $1+\bar{\sigma}$, and therefore increases when ponderable masses approach the test body.
2. There is an inductive action of accelerated masses, of the same sign, upon the test body. This is the term $\dfrac{\partial \mathfrak{A}}{\partial l}$.
3. A material particle, moving perpendicularly to the axis of rotation inside a rotating hollow body, is deflected in the sense of the rotation (Coriolis field). The centrifugal action, mentioned above, inside a rotating hollow body, also follows from the theory, as has been shown by Thirring.*

* That the centrifugal action must be inseparably connected with the existence of the Coriolis field may be recognized, even without calculation, in the special case of a co-ordinate system rotating uniformly relatively to an inertial system; our general co-variant equations naturally must apply to such a case.

Although all of these effects are inaccessible to experiment, because κ is so small, nevertheless they certainly exist according to the general theory of relativity. We must see in them a strong support for Mach's ideas as to the relativity of all inertial actions. If we think these ideas consistently through to the end we must expect the *whole* inertia, that is, the *whole* $g_{\mu\nu}$-field, to be determined by the matter of the universe, and not mainly by the boundary conditions at infinity.

For a satisfactory conception of the $g_{\mu\nu}$-field of cosmical dimensions, the fact seems to be of significance that the relative velocity of the stars is small compared to the velocity of light. It follows from this that, with a suitable choice of co-ordinates, g_{44} is nearly constant in the universe, at least, in that part of the universe in which there is matter. The assumption appears natural, moreover, that there are stars in all parts of the universe, so that we may well assume that the inconstancy of g_{44} depends only upon the circumstance that matter is not distributed continuously, but is concentrated in single celestial bodies and systems of bodies. If we are willing to ignore these more local non-uniformities of the density of matter and of the $g_{\mu\nu}$-field, in order to learn something of the geometrical properties of the universe as a whole, it appears natural to substitute for the actual distribution of masses a continuous distribution, and furthermore to assign to this distribution a uniform density σ. In this imagined universe all points with space directions will be geometrically equivalent; with respect to its space extension it will have a constant curvature, and will be cylindrical with respect to its x_4-co-ordinate. The possibility seems to be particularly satisfying that the universe is spatially bounded and thus, in accordance with our

assumption of the constancy of σ, is of constant curvature, being either spherical or elliptical; for then the boundary conditions at infinity which are so inconvenient from the standpoint of the general theory of relativity, may be replaced by the much more natural conditions for a closed space.

According to what has been said, we are to put

$$ds^2 = dx_4^2 - \gamma_{\mu\nu} dx_\mu dx_\nu \tag{119}$$

in which the indices μ and ν run from 1 to 3 only. The $\gamma_{\mu\nu}$ will be such functions of x_1, x_2, x_3 as correspond to a three-dimensional continuum of constant positive curvature. We must now investigate whether such an assumption can satisfy the field equations of gravitation.

In order to be able to investigate this, we must first find what differential conditions the three-dimensional manifold of constant curvature satisfies. A spherical manifold of three dimensions, embedded in a Euclidean continuum of four dimensions,* is given by the equations

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = a^2$$
$$dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 = ds^2.$$

By eliminating x_4, we get

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + \frac{(x_1 dx_1 + x_2 dx_2 + x_3 dx_3)^2}{a^2 - x_1^2 - x_2^2 - x_3^2}.$$

Neglecting terms of the third and higher degrees in the x_ν, we can put, in the neighbourhood of the origin of co-ordinates,

$$ds^2 = \left(\delta_{\mu\nu} + \frac{x_\mu x_\nu}{a^2}\right) dx_\mu dx_\nu.$$

* The aid of a fourth space dimension has naturally no significance except that of a mathematical artifice.

Inside the brackets are the $g_{\mu\nu}$ of the manifold in the neighbourhood of the origin. Since the first derivatives of the $g_{\mu\nu}$, and therefore also the $\Gamma^{\sigma}_{\mu\nu}$, vanish at the origin, the calculation of the $R_{\mu\nu}$ for this manifold, by (88), is very simple at the origin. We have

$$R_{\mu\nu} = -\frac{2}{a^2}\delta_{\mu\nu} = -\frac{2}{a^2}g_{\mu\nu}.$$

Since the relation $R_{\mu\nu} = -\frac{2}{a^2}g_{\mu\nu}$ is generally co-variant, and since all points of the manifold are geometrically equivalent, this relation holds for every system of co-ordinates, and everywhere in the manifold. In order to avoid confusion with the four-dimensional continuum, we shall, in the following, designate quantities that refer to the three-dimensional continuum by Greek letters, and put

$$\mathrm{P}_{\mu\nu} = -\frac{2}{a^2}\gamma_{\mu\nu}. \tag{120}$$

We now proceed to apply the field equations (96) to our special case. From (119) we get for the four-dimensional manifold,

$$\begin{cases} R_{\mu\nu} = \mathrm{P}_{\mu\nu} \text{ for the indices 1 to 3} \\ R_{14} = R_{24} = R_{34} = R_{44} = 0 \end{cases} \tag{121}$$

For the right-hand side of (96) we have to consider the energy tensor for matter distributed like a cloud of dust. According to what has gone before we must therefore put

$$T^{\mu\nu} = \sigma \frac{dx_\mu}{ds}\frac{dx_\nu}{ds}$$

specialized for the case of rest. But in addition, we shall add a pressure term that may be physically established as

follows. Matter consists of electrically charged particles. On the basis of Maxwell's theory these cannot be conceived of as electromagnetic fields free from singularities. In order to be consistent with the facts, it is necessary to introduce energy terms, not contained in Maxwell's theory, so that the single electric particles may hold together in spite of the mutual repulsions between their elements, charged with electricity of one sign. For the sake of consistency with this fact, Poincaré has assumed a pressure to exist inside these particles which balances the electrostatic repulsion. It cannot, however, be asserted that this pressure vanishes outside the particles. We shall be consistent with this circumstance if, in our phenomenological presentation, we add a pressure term. This must not, however, be confused with a hydrodynamical pressure, as it serves only for the energetic presentation of the dynamical relations inside matter. Accordingly we put

$$T_{\mu\nu} = g_{\mu\alpha}g_{\nu\beta}\sigma \frac{dx_\alpha}{ds}\frac{dx_\beta}{ds} - g_{\mu\nu}p. \tag{122}$$

In our special case we have, therefore, to put

$$\begin{aligned} T_{\mu\nu} &= \gamma_{\mu\nu}p \text{ (for } \mu \text{ and } \nu \text{ from 1 to 3)} \\ T_{44} &= \sigma - p \\ T &= -\gamma^{\mu\nu}\gamma_{\mu\nu}p + \sigma - p = \sigma - 4p. \end{aligned}$$

Observing that the field equation (96) may be written in the form

$$R_{\mu\nu} = -\kappa(T_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}T)$$

we get from (96) the equations,

$$\begin{aligned} +\frac{2}{a^2}\gamma_{\mu\nu} &= \kappa\left(\frac{\sigma}{2} - p\right)\gamma_{\mu\nu} \\ 0 &= -\kappa\left(\frac{\sigma}{2} + p\right). \end{aligned}$$

From this follows

$$\begin{cases} p = -\frac{\sigma}{2} \\ a = \sqrt{\frac{2}{\kappa\sigma}} \end{cases} \tag{123}$$

If the universe is quasi-Euclidean, and its radius of curvature therefore infinite, then σ would vanish. But it is improbable that the mean density of matter in the universe is actually zero; this is our third argument against the assumption that the universe is quasi-Euclidean. Nor does it seem possible that our hypothetical pressure can vanish; the physical nature of this pressure can be appreciated only after we have a better theoretical knowledge of the electromagnetic field. According to the second of equations (123) the radius, a, of the universe is determined in terms of the total mass, M, of matter, by the equation

$$a = \frac{M\kappa}{4\pi^2}. \tag{124}$$

The complete dependence of the geometrical upon the physical properties becomes clearly apparent by means of this equation.

Thus we may present the following arguments against the conception of a space-infinite, and for the conception of a space-bounded, or closed, universe:—

1. From the standpoint of the theory of relativity, to postulate a closed universe is very much simpler than to postulate the corresponding boundary condition at infinity of the quasi-Euclidean structure of the universe.

2. The idea that Mach expressed, that inertia depends upon the mutual action of bodies, is contained, to a first approximation, in the equations of the theory of relativity;

it follows from these equations that inertia depends, at least in part, upon mutual actions between masses. Thereby Mach's idea gains in probability, as it is an unsatisfactory assumption to make that inertia depends in part upon mutual actions, and in part upon an independent property of space. But this idea of Mach's corresponds only to a finite universe, bounded in space, and not to a quasi-Euclidean, infinite universe. From the standpoint of epistemology it is more satisfying to have the mechanical properties of space completely determined by matter, and this is the case only in a closed universe.

3. An infinite universe is possible only if the mean density of matter in the universe vanishes. Although such an assumption is logically possible, it is less probable than the assumption that there is a finite mean density of matter in the universe.

APPENDIX
FOR THE SECOND EDITION

ON THE "COSMOLOGIC PROBLEM"

SINCE the first edition of this little book some advances have been made in the theory of relativity. Some of these we shall mention here only briefly:

The first step forward is the conclusive demonstration of the existence of the red shift of the spectral lines by the (negative) gravitational potential of the place of origin (see p. 92). This demonstration was made possible by the discovery of so-called "dwarf stars" whose average density exceeds that of water by a factor of the order 10^4. For such a star (e.g. the faint companion of Sirius), whose mass and radius can be determined,* this red shift was expected, by the theory, to be about 20 times as large as for the sun, and indeed it was demonstrated to be within the expected range.

A second step forward, which will be mentioned briefly, concerns the law of motion of a gravitating body. In the initial formulation of the theory the law of motion for a gravitating particle was introduced as an independent fundamental assumption in addition to the field law of gravitation—see Eq. 90 which asserts that a gravitating particle moves in a geodesic line. This constitutes a

* The mass is derived from the reaction on Sirius by spectroscopic means, using the Newtonian laws; the radius is derived from the total lightness and from the intensity of radiation per unit area, which may be derived from the temperature of its radiation.

APPENDIX FOR THE SECOND EDITION

hypothetic translation of Galileo's law of inertia to the case of the existence of "genuine" gravitational fields. It has been shown that this law of motion—generalized to the case of arbitrarily large gravitating masses—can be derived from the field-equations of empty space alone. According to this derivation the law of motion is implied by the condition that the field be singular nowhere outside its generating mass points.

A third step forward, concerning the so-called "cosmologic problem," will be considered here in detail, in part because of its basic importance, partly also because the discussion of these questions is by no means concluded. I feel urged toward a more exact discussion also by the fact that I cannot escape the impression that in the present treatment of this problem the most important basic points of view are not sufficiently stressed.

The problem can be formulated roughly thus: On account of our observations on fixed stars we are sufficiently convinced that the system of fixed stars does not in the main resemble an island which floats in infinite empty space, and that there does not exist anything like a center of gravity of the total amount of existing matter. Rather, we feel urged toward the conviction that there exists an average density of matter in space which differs from zero.

Hence the question arises: Can this hypothesis, which is suggested by experience, be reconciled with the general theory of relativity?

First we have to formulate the problem more precisely. Let us consider a finite part of the universe which is large enough so that the average density of matter contained in it is an approximately continuous function of (x_1, x_2, x_3, x_4). Such a subspace can be considered approximately as an inertial system (Minkowski space) to which we relate the

APPENDIX FOR THE SECOND EDITION

motion of the stars. One can arrange it so that the mean velocity of matter relative to this system shall vanish in all directions. There remain the (almost random) motions of the individual stars, similar to the motions of the molecules of a gas. It is essential that the velocities of the stars are known by experience to be very small as compared to the velocity of light. It is therefore feasible for the moment to neglect this relative motion completely, and to consider the stars replaced by material dust without (random) motion of the particles against each other.

The above conditions are by no means sufficient to make the problem a definite one. The simplest and most radical specialization would be the condition: The (naturally measured) density, ρ of matter is the same everywhere in (four-dimensional) space, the metric is, for a suitable choice of coordinates, independent of x_4 and homogeneous and isotropic with respect to x_1, x_2, x_3.

It is this case which I at first considered the most natural idealized description of physical space in the large; it is treated on pages 103–108 of this book. The objection to this solution is that one has to introduce a negative pressure, for which there exists no physical justification. In order to make that solution possible I originally introduced a new member into the equation instead of the above mentioned pressure, which is permissible from the point of view of relativity. The equations of gravitation thus enlarged were:

$$(1) \qquad (R_{ik} - \tfrac{1}{2}g_{ik}R) + \Lambda g_{ik} + \kappa T_{ik} = 0$$

where Λ is a universal constant ("cosmologic constant"). The introduction of this second member constitutes a complication of the theory, which seriously reduces its logical simplicity. Its introduction can only be justified

APPENDIX FOR THE SECOND EDITION

by the difficulty produced by the almost unavoidable introduction of a finite average density of matter. We may remark, by the way, that in Newton's theory there exists the same difficulty.

The mathematician Friedman found a way out of this dilemma.* His result then found a surprising confirmation by Hubble's discovery of the expansion of the stellar system (a red shift of the spectral lines which increases uniformly with distance). The following is essentially nothing but an exposition of Friedman's idea:

FOUR-DIMENSIONAL SPACE WHICH IS ISOTROPIC WITH RESPECT TO THREE DIMENSIONS

We observe that the systems of stars, as seen by us, are spaced with approximately the same density in all directions. Thereby we are moved to the assumption that the *spatial* isotropy of the system would hold for all observers, for every place and every time of an observer who is at rest as compared with surrounding matter. On the other hand we no longer make the assumption that the average density of matter, for an observer who is at rest relative to neighboring matter, is constant with respect to time. With this we drop the assumption that the expression of the metric field is independent of time.

We now have to find a mathematical form for the condition that the universe, *spatially speaking*, is isotropic everywhere. Through every point P of (four-dimensional) space there is the path of a particle (which in the following will be called "geodesic" for short). Let P and Q be two

* He showed that it is possible, according to the field equations, to have a finite density in the whole (three-dimensional) space, without enlarging these field equations *ad hoc*. Zeitschr. f. Phys. 10 (1922).

APPENDIX FOR THE SECOND EDITION

infinitesimally near points of such a geodesic. We shall then have to demand that the expression of the field shall be invariant relative to any rotation of the coordinate system keeping P and Q fixed. This will be valid for any element of any geodesic.*

The condition of the above invariance implies that the entire geodesic lies on the axis of rotation and that its points remain invariant under rotation of the coordinate system. This means that the solution shall be invariant with respect to all rotations of the coordinate system around the triple infinity of geodesics.

For the sake of brevity I will not go into the deductive derivation of the solution of this problem. It seems intuitively evident, however, for the three-dimensional space that a metric which is invariant under rotations around a double infinity of lines will be essentially of the type of central symmetry (by suitable choice of coordinates), where the axes of rotations are the radial straight lines, which by reasons of symmetry are geodesics. The surfaces of constant radius are then surfaces of constant (positive) curvature which are everywhere perpendicular to the (radial) geodesics. Hence we obtain in invariant language:

There exists a family of surfaces orthogonal to the geodesics. Each of these surfaces is a surface of constant curvature. The segments of these geodesics contained between any two surfaces of the family are equal.

Remark. The case which has thus been obtained intuitively is not the general one in so far as the surfaces of the family could be of constant negative curvature or Euclidean (zero curvature).

* This condition not only limits the metric, but it necessitates that for every geodesic there exist a system of coordinates such that relative to this system the invariance under rotation around this geodesic is valid.

APPENDIX FOR THE SECOND EDITION

The four-dimensional case which interests us is entirely analogous. Furthermore there is no essential difference when the metric space is of index of inertia 1; only that one has to choose the radial directions as timelike and correspondingly the directions in the surfaces of the family as spacelike. The axes of the local light cones of all points lie on the radial lines.

CHOICE OF COORDINATES

Instead of the four coordinates for which the spatial isotropy of the universe is most clearly apperent, we now choose different coordinates which are more convenient from the point of view of physical interpretation.

As timelike lines on which x_1, x_2, x_3 are constant and x_4 alone variable we choose the particle geodesics which in the central symmetric form are the straight lines through the center. Let x_4 further equal the metric distance from the center. In such coordinates the metric is of the form:

$$(2) \qquad \begin{cases} ds^2 = dx_4^2 - d\sigma^2 \\ d\sigma^2 = \gamma_{ik} dx_i dx_k \qquad (i, k = 1, 2, 3) \end{cases}$$

$d\sigma^2$ is the metric on one of the spherical hypersurfaces. The γ_{ik} which belong to different hypersurfaces will then (because of the central symmetry) be the same form on all hypersurfaces except for a positive factor which depends on x_4 alone:

$$(2a) \qquad \gamma_{ik} = \underset{0}{\gamma_{ik}} G^2$$

where the $\underset{0}{\gamma}$ depend on x_1, x_2, x_3 only, and G is a function of x_4 alone. We have then:

$$(2b) \qquad \underset{0}{d\sigma^2} = \underset{0}{\gamma_{ik}} dx_i dx_k \qquad (i, k = 1, 2, 3)$$

APPENDIX FOR THE SECOND EDITION

is a definite metric of constant curvature in three dimensions, the same for every G.

Such a metric is characterized by the equations:

$$\underset{0}{R}_{iklm} - B(\underset{0}{\gamma}_{il}\underset{0}{\gamma}_{km} - \underset{0}{\gamma}_{im}\underset{0}{\gamma}_{kl}) = 0 \tag{2c}$$

We can choose the coordinate system (x_1, x_2, x_3) so that the line element becomes conformally Euclidean:

$$\underset{0}{d\sigma}^2 = A^2(dx_1^2 + dx_2^2 + dx_3^2) \text{ i.e. } \underset{0}{\gamma}_{ik} = A^2\delta_{ik} \tag{2d}$$

where A shall be a positive function of $r(r = x_1^2 + x_2^2 + x_3^2)$ alone. By substitution into the equations, we get for A the two equations:

$$\begin{cases} -\frac{1}{r}\left(\frac{A'}{Ar}\right)' + \left(\frac{A'}{Ar}\right)^2 = 0 \\ -\frac{2A'}{Ar} - \left(\frac{A'}{A}\right)^2 - BA^2 = 0 \end{cases} \tag{3}$$

The first equation is satisfied by:

$$A = \frac{c_1}{c_2 + c_3 r^2} \tag{3a}$$

where the constants are arbitrary for the time being. The second equation then yields:

$$B = 4\frac{c_2 c_3}{c_1^2} \tag{3b}$$

About the constants c we get the following: If for $r = 0$, A shall be positive, then c_1 and c_2 must have the same sign. Since a change of sign of all three constants does not change A, we can make c_1 and c_2 both positive. We can also make c_2 equal to 1. Furthermore, since a positive factor

APPENDIX FOR THE SECOND EDITION

can always be incorporated into the G^2, we can also make c_1 equal to 1 without loss of generality. Hence we can set:

$$(3c) \qquad A = \frac{1}{1 + cr^2};\ B = 4c$$

There are now three cases:

$c > 0$ (spherical space)
$c < 0$ (pseudospherical space)
$c = 0$ (Euclidean space)

By a similarity transformation of coordinates ($x_1' = ax_i$, where a is constant), we can further get in the first case $c = \frac{1}{4}$, in the second case $c = -\frac{1}{4}$.

For the three cases we then have respectively:

$$(3d) \qquad \begin{cases} A = \dfrac{1}{1 + \dfrac{r^2}{4}};\ B = +1 \\ A = \dfrac{1}{1 - \dfrac{r^2}{4}};\ B = -1 \\ A = 1;\ B = 0 \end{cases}$$

In the spherical case the "circumference" of the unit space ($G = 1$) is $\int_{-\infty}^{\infty} \frac{dr}{1 + \frac{r^2}{4}} = 2\pi$, the "radius" of the unit space is 1. In all three cases the function G of time is a measure for the change with time of the distance of two points of matter (measured on a spatial section). In the spherical case, G is the radius of space at the time x_4.

Summary. The hypothesis of *spatial* isotropy of our idealized universe leads to the metric:

APPENDIX FOR THE SECOND EDITION

(2) $$ds^2 = dx_4^2 - G^2A^2(dx_1^2 + dx_2^2 + dx_3^2)$$

where G depends on x_4 alone, A on r $(= x_1^2 + x_2^2 + x_3^2)$ alone, where:

(3) $$A = \frac{1}{1 + \frac{z}{4}r^2}$$

and the different cases are characterized by $z = 1$, $z = -1$, and $z = 0$ respectively.

THE FIELD EQUATIONS

We must now further satisfy the field equations of gravitation, that is to say the field equations without the "cosmologic member" which had been introduced previously *ad hoc:*

(4) $$(R_{ik} - \tfrac{1}{2}g_{ik}R) + \kappa T_{ik} = 0$$

By substitution of the expression for the metric, which was based on the assumption of spatial isotropy, we get after calculation:

$$R_{ik} - \frac{1}{2}g_{ik}R = \left(\frac{z}{G^2} + \frac{G'^2}{G^2} + 2\frac{G''}{G}\right)GA\delta_{ik} \qquad (i, k = 1, 2, 3)$$

(4a) $$R_{44} - \frac{1}{2}g_{44}R = -3\left(\frac{z}{G^2} + \frac{G'^2}{G^2}\right)$$

$$R_{i4} - \tfrac{1}{2}g_{i4}R = 0 \qquad (i = 1, 2, 3)$$

Further we have for T_{ik}, the energy tensor of matter, for "dust":

(4b) $$T^{ik} = \rho\frac{dx_i}{ds}\frac{dx_k}{ds}$$

APPENDIX FOR THE SECOND EDITION

The geodesics, along which the matter moves, are the lines along which x_4 alone varies; on them $dx_4 = ds$. We have:

(4c) $$T^{44} = \rho$$

the only component different from zero. By lowering of the indices we get as the only non-vanishing component of T_{ik}:

(4d) $$T_{44} = \rho$$

Considering this, the field equations are:

(5) $$\begin{cases} \frac{z}{G^2} + \frac{G'^2}{G^2} + 2\frac{G''}{G} = 0 \\ \frac{z}{G^2} + \frac{G'^2}{G^2} - \frac{1}{3}\kappa\rho = 0 \end{cases}$$

$\frac{z}{G^2}$ is the curvature in the spatial section $x_4 = \text{const.}$ Since G is in all cases a relative measure for the metric distance of two material particles as function of time, $\frac{G'}{G}$ expresses Hubble's expansion. A drops out of the equations, as it has to if there shall be solutions of the equations of gravity of the required symmetrical type. By subtraction of both equations we get:

(5a) $$\frac{G''}{G} + \frac{1}{6}\kappa\rho = 0$$

Since G and ρ must be everywhere positive, G'' is everywhere negative for nonvanishing ρ. $G(x_4)$ can thus have no minimum nor a point of inflection; further there is no solution for which G is constant.

APPENDIX FOR THE SECOND EDITION

THE SPECIAL CASE OF VANISHING SPATIAL CURVATURE ($z = 0$)

The simplest special case for non-vanishing density ρ is the case $z = 0$, where the sections $x_4 = \text{const}$ are not curved. If we set $\frac{G'}{G} = h$, the field equations in this case are:

$$\text{(5b)} \qquad \begin{cases} 2h' + 3h^2 = 0 \\ \qquad 3h^2 = \kappa\rho \end{cases}$$

The relation between Hubble's expansion h and the average density ρ, which is given in the second equation, is comparable to some extent with experience, at least as far as the order of magnitude is concerned. The expansion is given as 432 km/sec for the distance of 10^6 parsec. If we express this in the system of measures used by us (cm—as unit length; unit of time—that of motion of light of one cm) we get:

$$h = \frac{432 \cdot 10^5}{3.25 \cdot 10^6 \cdot 365 \cdot 24 \cdot 60 \cdot 60} \cdot \left(\frac{1}{3 \cdot 10^{10}}\right)^2 = 4.71 \cdot 10^{-28}.$$

Since further (see 105a) $\kappa = 1.86\ 10^{-27}$, the second equation of (5b) yields:

$$\rho = \frac{3h^2}{\kappa} = 3.5 \cdot 10^{-28}\ \text{g./cm.}^3$$

This value corresponds, according to the order of magnitude, somewhat with the estimates given by astronomers (on the basis of the masses and parallaxes of visible stars and systems of stars). I quote here as example G. C. McVittie (Proceedings of the Physical Society of London, vol. 51, 1939, p. 537): "The average density is certainly not

APPENDIX FOR THE SECOND EDITION

greater than 10^{-27} g./cm.^{3} and is more probably of the order 10^{-29} g./cm.^{3}"

Owing to the great difficulty of determining this magnitude I consider this for the time being a satisfactory correspondence. Since the quantity h is determined with greater accuracy than ρ, it is probably not an exaggeration to assert that the determination of the structure of observable space is tied up with the more precise determination of ρ. Because, due to the second equation of (5), the space curvature is given in the general case as:

$$zG^{-2} = \tfrac{1}{3}\kappa\rho - h^2. \tag{5c}$$

Hence, if the right side of the equation is positive, the space is of positive constant curvature and therefor finite; its magnitude can be determined with the same accuracy as this difference. If the right side is negative, the space is infinite. At present ρ is not sufficiently determined to enable us to deduce from this relation a non-vanishing mean curvature of space (the section $x_4 = \text{const}$).

In case we neglect spatial curvature, the first equation of (5c) becomes, after suitable choice of the initial point of x_4:

$$h = \frac{2}{3} \cdot \frac{1}{x_4} \tag{6}$$

This equation has a singularity for $x_4 = 0$, so that such a space has either a negative expansion and the time is limited from above by the value $x_4 = 0$, or it has a positive expansion and begins to exist for $x_4 = 0$. The latter case corresponds to what we find realized in nature.

From the measured value of h we get for the time of existence of the world up to now $1.5 \cdot 10^9$ years. This age is about the same as that which one has obtained from the

APPENDIX FOR THE SECOND EDITION

disintegration of uranium for the firm crust of the earth. This is a paradoxical result, which for more than one reason has aroused doubts as to the validity of the theory.

The question arises: Can the present difficulty, which arose under the assumption of a practically negligible spatial curvature, be eliminated by the introduction of a suitable spatial curvature? Here the first equation of (5), which determines the time-dependence of G, will be of use.

SOLUTION OF THE EQUATIONS IN THE CASE OF NON-VANISHING SPATIAL CURVATURE

If one considers a spatial curvature of the spatial section ($x_4 = \text{const}$), one has the equations:

$$\text{(5)} \qquad \begin{aligned} zG^{-2} + \left(2\frac{G''}{G} + \left(\frac{G'}{G}\right)^2\right) &= 0 \\ zG^{-2} + \left(\frac{G'}{G}\right)^2 - \frac{1}{3}\kappa\rho &= 0 \end{aligned}$$

The curvature is positive for $z = +1$, negative for $z = -1$. The first of these equations is integrable. We first write it it in the form:

$$\text{(5d)} \qquad z + 2GG'' + G'^2 = 0.$$

If we consider x_4 $(= t)$ as a function of G, we have:

$$G' = \frac{1}{t'},\ G'' = \left(\frac{1}{t'}\right)' \frac{1}{t'}.$$

If we write $u(G)$ for $\frac{1}{t'}$, we get:

$$\text{(5e)} \qquad z + 2Guu' + u^2 = 0$$

or

$$\text{(5f)} \qquad z + (Gu^2)' = 0.$$

APPENDIX FOR THE SECOND EDITION

From this we get by simple integration:

(5g) $$zG + Gu^2 = G_0$$

or, since we set $u = \dfrac{1}{\dfrac{dt}{dG}} = \dfrac{dG}{dt}$:

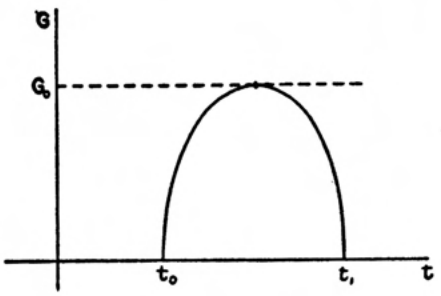
(5h) $$\left(\frac{dG}{dt}\right)^2 = \frac{G_0 - zG}{G}$$

where G_0 is a constant. This constant cannot be negative, as we see if we differentiate (5h) and consider that G'' is negative because of (5a).

(a) Space with positive curvature

G remains in the interval $0 \leq G \leq G_0$. G is given quantitatively by a sketch like the following:

(1)



The radius G rises from 0 to G_0 and then again drops continuously to 0. The spatial section is finite (spherical)

(5c) $$\tfrac{1}{3}\kappa\rho - h^2 > 0$$

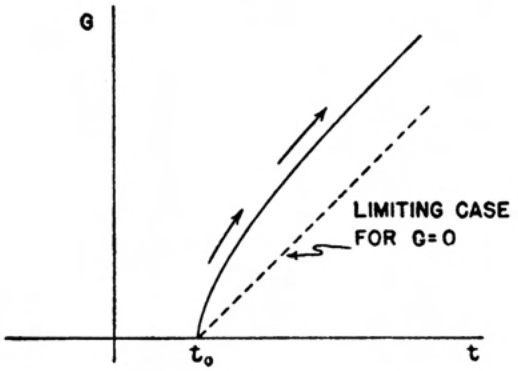
(b) Space with negative curvature

$$\left(\frac{dG}{dt}\right)^2 = \frac{G_0 + G}{G}.$$

APPENDIX FOR THE SECOND EDITION

G increases with t from $G = 0$ to $G = +\infty$ (or goes from $G = \infty$ to $G = 0$). Hence $\frac{dG}{dt}$ decreases monotonically from $+\infty$ to 1 as illustrated by the sketch:

(2)



This is then a case of continued expansion with no contraction. The spatial section is infinite, and we have:

(5c) $$\tfrac{1}{3}\kappa\rho - h^2 < 0.$$

The case of plane spatial section, which was treated in the previous section, lies between these two cases, according to the equation:

(5h) $$\left(\frac{dG}{dt}\right)^2 = \frac{G_0}{G}.$$

Remark. The case of negative curvature contains as a limiting case that of vanishing ρ. For this case $\left(\frac{dG}{dt}\right)^2 = 1$ (see sketch 2). This is the Euclidean case; since the calculations show that the curvature tensor vanishes.

The case of negative curvature with non-vanishing ρ approaches this limiting case more and more closely, so

APPENDIX FOR THE SECOND EDITION

that with increasing time the structure of space will be less and less determined by the matter contained in it.

From this investigation of the case of non-vanishing curvature results the following. For every state of non-vanishing ("spatial") curvature, there exists, as in the case of vanishing curvature, an initial state in which $G = 0$ where the expansion starts. Hence this is a section at which the density is infinite and the field is singular. The introduction of such a new singularity seems problematical in itself.*

It appears, further, that the influence of the introduction of a spatial curvature on the time interval between the start of the expansion and the drop to a fixed value $h = \frac{G'}{G}$ is negligible in its order of magnitude. This time interval may be obtained by elementary calculations from (5h), which we shall omit here. We restrict ourselves to the consideration of an expanding space with vanishing ρ. This, as mentioned before, is a special case of negative spatial curvature. The second equation of (5) yields (considering the reversal of sign of the first member):

$$G' = 1.$$

Hence (for suitable initial point for x_4)

$$G = x_4$$

(6a) $$h = \frac{G'}{G} = \frac{1}{x_4} \cdots$$

Hence this extreme case yields for the duration of the expansion the same result as the case of vanishing spatial

* However, the following should be noted: The present relativistic theory of gravitation is based on a separation of the concepts of "gravitational field" and of "matter." It may be plausible that the theory is for this reason inadequate for very high density of matter. It may well be the case that for a unified theory there would arise no singularity.

APPENDIX FOR THE SECOND EDITION

curvature (see Eq. 6) except for a factor of order of magnitude 1.

The doubts mentioned in connection with Eq. (6), namely that this would give such a remarkably short duration for the development of the stars and systems of stars which are observable at present, cannot therefore be removed by the introduction of a spatial curvature.

EXTENSION OF THE ABOVE CONSIDERATIONS BY GENERALIZATION OF THE EQUATION WITH RESPECT TO PONDERABLE MATTER

For all the solutions reached up to now, there exists a state of the system at which the metric becomes singular ($G = 0$) and the density ρ becomes infinite. The following question arises: Could not the rise of such singularities be due to the fact that we introduced matter as a kind of dust which does not resist condensation? Did we not neglect, without justification, the influence of the random motion of the single stars?

One could, for example, replace dust whose particles are at rest relative to each other, by one whose particles are in random motion relative to each other like the molecules of a gas. Such matter would offer a resistance to adiabatic condensation which increases with that condensation. Will not this be able to prevent the rise of infinite condensation? We shall show below that such a modification in the description of matter can change nothing of the main character of the above solutions.

"PARTICLE-GAS" TREATED ACCORDING TO SPECIAL RELATIVITY

We consider a swarm of particles of mass m in parallel motion. By a proper transformation this swarm can be

APPENDIX FOR THE SECOND EDITION

considered at rest. The spatial density of the particles, σ, is then invariant in the Lorentz sense. Related to an arbitrary Lorentz system

$$T^{uv} = m\sigma \frac{dx^u}{ds}\frac{dx^v}{ds} \tag{7}$$

has invariant meaning (energy tensor of the swarm). If there exist many such swarms we get, by summation, for all of them:

$$T^{uv} = m \sum_p \sigma_p \left(\frac{dx^u}{ds}\right)_p \left(\frac{dx^v}{ds}\right)_p \tag{7a}$$

In relation to this form we can choose the time axis of the Lorentz system so that: $T^{14} = T^{24} = T^{34} = 0$. Further, we can obtain by spatial rotation of the system: $T^{12} = T^{23} = T^{31} = 0$. Let, further, the particle gas be isotropic. This means that $T^{11} = T^{22} = T^{33} = p$. This is an invariant as well as $T^{44} = u$. The invariant:

$$\mathcal{J} = T^{uv}g_{uv} = T^{44} - (T^{11} + T^{22} + T^{33}) = u - 3p \tag{7b}$$

is thus expressed in terms of u and p.

It follows from the expression for T^{uv} that T^{11}, T^{22}, T^{33} and T^{44} are all positive; hence the same is true for T_{11}, T_{22}, T_{33}, T_{44}.

The equations of gravity are now:

$$\begin{aligned} 1 + 2GG'' + G^2 + \kappa T_{11} &= 0 \\ -3G^{-2}(1 + G'^2) + \kappa T_{44} &= 0. \end{aligned} \tag{8}$$

From the first it follows that here too (since $T_{11} > 0$) G'' is always negative where the member T_{11} for given G and G' can only decrease the value of G''.

APPENDIX FOR THE SECOND EDITION

From this we see that the consideration of a random relative motion of the mass points does not change our results fundamentally.

SUMMARY AND OTHER REMARKS

(1) The introduction of the "cosmologic member" into the equations of gravity, though possible from the point of view of relativity, is to be rejected from the point of view of logical economy. As Friedman was the first to show one can reconcile an everywhere finite density of matter with the original form of the equations of gravity if one admits the time variability of the metric distance of two mass points.*

(2) The demand for *spatial* isotropy of the universe alone leads to Friedman's form. It is therefore undoubtedly the general form, which fits the cosmologic problem.

(3) Neglecting the influence of spatial curvature, one obtains a relation between the mean density and Hubble's expansion which, as to order of magnitude, is confirmed empirically.

One further obtains, for the time from the start of the expansion up to the present, a value of the order of magnitude of 10^9 years. The brevity of this time does not concur with the theories on the developments of fixed stars.

(4) The latter result is not changed by the introduction of spatial curvature; nor is it changed by the consideration of the random motion of stars and systems of stars with respect to each other.

* If Hubble's expansion had been discovered at the time of the creation of the general theory of relativity, the cosmologic member would never have been introduced. It seems now so much less justified to introduce such a member into the field equations, since its introduction loses its sole original justification,—that of leading to a natural solution of the cosmologic problem.

APPENDIX FOR THE SECOND EDITION

(5) Some try to explain Hubble's shift of spectral lines by means other than the Doppler effect. There is, however, no support for such a conception in the known physical facts. According to such a hypothesis it would be possible to connect two stars, S_1 and S_2, by a rigid rod. Monochromatic light which is sent from S_1 to S_2 and reflected back to S_1 could arrive with a different frequency (measured by a clock on S_1) if the number of wave lengths of light along the rod should change with time on the way. This would mean that the locally measured velocity of light would depend on time, which would contradict even the special theory of relativity. Further it should be noted that a light signal going to and fro between S_1 and S_2 would constitute a "clock" which would not be in a constant relation with a clock (e.g. an atomistic clock) in S_1. This would mean that there would exist no metric in the sense of relativity. This not only involves the loss of comprehension of all those relations which relativity has yielded, but it also fails to concur with the fact that certain atomistic forms are not related by "similarity" but by "congruence" (the existence of sharp spectral lines, volumes of atoms, etc.).

The above considerations are, however, based on wave theory, and it may be that some proponents of the above hypothesis imagine that the process of the expansion of light is altogether not according to wave theory, but rather in a manner analogous to the Compton effect. The assumption of such a process without scattering constitutes a hypothesis which is not justified from the point of view of our present knowledge. It also fails to give a reason for the independence of the relative shift of frequence from the original frequency. Hence one cannot but consider Hubble's discovery as an expansion of the system of stars.

APPENDIX FOR THE SECOND EDITION

(6) The doubts about the assumption of a "beginning of the world" (start of the expansion) only about 10^9 years ago have roots of both an empirical and a theoretical nature. The astronomers tend to consider the stars of different spectral types as age classes of a uniform development, which process would need much longer than 10^9 years. Such a theory therefore actually contradicts the demonstrated consequences of the relativistic equations. It seems to me, however, that the "theory of evolution" of the stars rests on weaker foundations than the field equations.

The theoretical doubts are based on the fact that for the time of the beginning of the expansion the metric becomes singular and the density, ρ, becomes infinite. In this connection the following should be noted: The present theory of relativity is based on a division of physical reality into a metric field (gravitation) on the one hand, and into an electromagnetic field and matter on the other hand. In reality space will probably be of a uniform character and the present theory be valid only as a limiting case. For large densities of field and of matter, the field equations and even the field variables which enter into them will have no real significance. One may not therefore assume the validity of the equations for very high density of field and of matter, and one may not conclude that the "beginning of the expansion" must mean a singularity in the mathematical sense. All we have to realize is that the equations may not be continued over such regions.

This consideration does, however, not alter the fact that the "beginning of the world" really constitutes a beginning, from the point of view of the development of the now existing stars and systems of stars, at which those stars and systems of stars did not yet exist as individual entities.

APPENDIX FOR THE SECOND EDITION

(7) There are, however, some empirical arguments in *favor* of a dynamic concept of space as required by the theory. Why does there still exist uranium, despite its comparatively rapid decomposition, and despite the fact that no possibility for the creation of uranium is recognizable? Why is space not so filled with radiation as to make the nocturnal sky look like a glowing surface? This is an old question which so far has found no satisfactory answer from the point of view of a stationary world. But it would lead too far to go into questions of this type.

(8) For the reasons given it seems that we have to take the idea of an expanding universe seriously, in spite of the short "lifetime." If one does so, the main question becomes whether space has positive or negative spatial curvature. To this we add the following remark.

From the empirical point of view the decision boils down to the question whether the expression $\frac{1}{3}\kappa\rho - h^2$ is positive (spherical case) or negative (pseudospherical case). This seems to me to be the most important question. An empirical decision does not seem impossible at the present state of astronomy. Since h (Hubble's expansion) is comparatively well known, everything depends on determining ρ with the highest possible accuracy.

It is imaginable that the proof would be given that the world is spherical (it is hardly imaginable that one could prove it to be pseudospherical). This depends on the fact that one can always give a lower bound for ρ but not an upper bound. This is the case because we can hardly form an opinion on how large a fraction of ρ is given by astronomically unobservable (not radiating) masses. This I wish to discuss in somewhat greater detail.

One can give a lower bound for ρ (ρ_s) by taking into consideration only the masses of radiating stars. If it

APPENDIX FOR THE SECOND EDITION

should appear that $\rho_s > \frac{3h^2}{\kappa}$ then one would have decided in favor of spherical space. If it appears that $\rho_s < \frac{3h^2}{\kappa}$ one has to try to determine the share of non-radiating masses ρ_d. We want to show that one can also find a lower bound for $\frac{\rho_d}{\rho_s}$.

We consider an astronomical object which contains many single stars and which can be considered with sufficient accuracy to be a stationary system, e.g. a globular cluster (of known parallax). From the velocities which are observable spectroscopically one can determine the field of gravitation (under plausible assumptions) and thereby the masses which generate this field. The masses which are so computed one can compare with those of the visible stars of the cluster, and so find at least a rough approximation for how far the masses which generate the field exceed those of the visible stars of the cluster. One obtains thus an estimate for $\frac{\rho_d}{\rho_s}$ for the particular cluster.

Since the non-radiating stars will on the average be smaller than the radiating ones, they will tend on the average to greater velocities than the larger stars due to their interaction with the stars of the cluster. Hence they will "evaporate" more quickly from the cluster than the larger stars. It may therefore be expected that the relative frequency of the smaller heavenly bodies inside the cluster will be smaller than that outside of it. One can therefore obtain in $\left(\frac{\rho_d}{\rho_s}\right)_k$ (relation of densities in the above cluster) a lower bound for the ratio $\frac{\rho_d}{\rho_s}$ in the whole

APPENDIX FOR THE SECOND EDITION

space. One therefore obtains as a lower bound for the entire average density of mass in space:

$$\rho_s \left[1 + \left(\frac{\rho_d}{\rho_s} \right)_k \right]$$

If this quantity is greater than $\frac{3h^2}{\kappa}$ one may conclude that space is of a spherical character. On the other hand I cannot think of any reasonably reliable determination of an upper bound for ρ.

(9) Last and not least: The age of the universe, in the sense used here, must certainly exceed that of the firm crust of the earth as found from the radioactive minerals. Since determination of age by these minerals is reliable in every respect, the cosmologic theory here presented would be disproved if it were found to contradict any such results. In this case I see no reasonable solution.

APPENDIX II

GENERALIZATION OF GRAVITATION THEORY

THE content of the general theory of relativity presented above is formally expressed by the equation

$$R_{ik} - \tfrac{1}{2}g_{ik}R = T_{ik} . \tag{1}$$

The left-hand side of this equation depends only on the symmetrical tensor, g_{ik}, which represents the metric properties of space as well as the gravitational field. The right-hand side of (1) is a phenomenological description of all the sources of the gravitational field. T_{ik} represents the energy which generates the gravitational field, but is itself of non-gravitational character, as for example the energy of the electromagnetic field, of the density of ponderable matter etc. To represent this tensor, T_{ik}, concepts are taken from pre-relativity physics which only a posteriori have been adapted to the principle of general relativity.

Such a dualistic treatment of the total field could not be avoided during the first phase of the development of relativity. It was, however, evident that this represented only a provisional treatment of the total problem. The starting point of the theory was the recognition of the unity of gravitation and inertia (principle of equivalence). From this principle, in conjunction with the knowledge that light exhibits a definite behavior in "empty space," it followed that the properties of the latter were to be represented by a symmetrical tensor, g_{ik}. The principle of equivalence, however, does not give any clue as to what may be the more comprehensive mathematical structure on which to base the treatment of the total field comprising the entire physical reality. In this case, our knowledge of physics does not

APPENDIX II

permit of such an unequivocal choice as does the equivalence of inertia and gravitation with respect to the partial field of pure gravitation. The only clue which can be drawn from experience is the vague perception that something like Maxwell's electromagnetic field has to be contained within the total field.

As a result, the attempt at a consistent extension of the theory of general relativity leads to an apparently quite vague formulation of the problem: to find a field structure which is a natural generalization of the symmetrical tensor as well as a system of field equations for this structure which represent a natural generalization of the equations

$$R_{ik} = 0 \tag{2}$$

of pure gravitation.

In the following, I shall present an attempt at the solution of this problem which appears to me highly convincing although, due to mathematical difficulties, I have not yet found a practicable way to confront the results of the theory with experimental evidence.

§1. THE FIELD STRUCTURE

The first problem is: how can the field structure be generalized in a natural way? Which is the natural generalization of the symmetrical tensor g_{ik} ?

The solution which comes immediately to mind is the substitution of the symmetrical tensor field by a non-symmetrical one. It is, indeed, this field structure which is chosen as the basis of the theory to be presented here.

But it is easily understood why this choice which at first appears so natural has, until lately, never been seriously considered. If the tensor g_{ik} is decomposed into its symmetrical and skew-symmetrical parts ($g_{\underline{ik}}$ and $g_{\underset{\vee}{ik}}$ respectively) according to the equation

$$g_{ik} = g_{\underline{ik}} + g_{\underset{\vee}{ik}} \tag{3}$$

then each of these two parts is by itself a tensor, i.e. under a co-

APPENDIX II

ordinate transformation, the components of each part transform independently of the components of the other part. In other words, considered from the point of view of the relativistic group, the non-symmetrical g_{ik} is not an irreducible entity but an arbitrary and unjustified combination of two entities of different nature.

This would seem to be a grave objection. It should be noted, however, that the group-theoretical point of view is not the only relevant one from which to judge the "uniformity" of the concept of the non-symmetrical tensor field. In fact, in the theory of the symmetrical tensor field (Riemann's theory), the determinant of the g_{ik} is a quantity of great importance. In particular, it makes it possible to correlate a contra-variant g^{ik} to the covariant tensor g_{ik} according to the equation

$$g_{is}\, g^{it} = \delta_s^t = g_{si}\, g^{ti} \tag{4}$$

where δ_s^t is the Kronecker tensor. This correlation which plays a fundamental part in the theory of the symmetrical field can immediately be taken over in the case of the non-symmetrical field; in the latter case, however, the order of the indices must be preserved. This represents one argument indicating that, in spite of the objection expressed above, it is indeed natural to introduce the non-symmetrical tensor field as a generalization of the symmetrical one. Equation (4) makes it possible to raise and lower the tensor indices; in the case of the non-symmetrical field, however, this operation is no longer uniquely defined a priori ($A_s\, g^{sk}$ and $A_s g^{ks}$ are not equal to each other) even though the "fundamental tensor" is given.

§2. AFFINE DISPLACEMENT AND ABSOLUTE DIFFERENTIAL CALCULUS IN THE CASE OF THE NON-SYMMETRICAL FIELD

The notion of the infinitesimal parallel displacement of a vector

$$\delta A^i = -\Gamma_{st}^i\, A^s\, dx^t \tag{5a}$$

$$\delta A_i = \Gamma_{it}^s\, A_s\, dx^t \tag{5b}$$

APPENDIX II

can immediately be extended to our generalized theory. But it proves then necessary to drop the symmetry of Γ in the two lower indices so that a reasonable relation between the g_{ik} and the Γ^l_{ik} may be postulated.

From (5a) and (5b) it follows that*

(6a) $$A^i{}_{,k} + A^s\Gamma^i_{sk}$$

and

(6b) $$A_{i,k} - A_s\Gamma^s_{ik}$$

have tensor character, just as in the case of the symmetrical Γ^l_{ik}. The definitions (5) imply the transformation law for the Γ^l_{ik}, just as they do for the symmetrical Γ:

$$(7) \qquad \Gamma^{l^*}_{ik} = \frac{\partial x^{l^*}}{\partial x^l} \cdot \frac{\partial x^r}{\partial x^{i^*}} \cdot \frac{\partial x^s}{\partial x^{k^*}} \Gamma^l_{rs} - \frac{\partial^2 x^{l^*}}{\partial x^a \partial x^b} \cdot \frac{\partial x^a}{\partial x^{i^*}} \cdot \frac{\partial x^b}{\partial x^{k^*}}.$$

One can easily infer that by decomposing the Γ^l_{ik} according to the symmetry of the lower indices,

(8) $$\Gamma^l_{ik} = \Gamma^l_{\underline{ik}} + \Gamma^l_{\underset{\vee}{ik}}$$

two quantities are formed which transform independently of each other. By interchanging i and k in (6b) and subtracting the tensor so obtained from (6b) we get the tensor

$$(A_{i,k} - A_{k,i}) - A_s(\Gamma^s_{ik} - \Gamma^s_{ki}).$$

Since the first parenthesis has tensor character, the same is true for $A_s\Gamma^s_{\underset{\vee}{ik}}$, hence also for $\Gamma^s_{\underset{\vee}{ik}}$. It then follows from (6b) that

$$A_{i,k} - A_s\Gamma^s_{\underline{ik}}$$

is also a tensor. Consequently, $\Gamma^s_{\underline{ik}}$ is itself a (symmetrical) displacement field.

*In the following the comma will always denote partial differentiation; thus, for example, $A^i{}_{,k} \equiv \frac{\partial A^i}{\partial x^k}$.

APPENDIX II

From a purely group-theoretical point of view, the Γ^l_{ik}, too, do not form a uniform entity since the decomposition according to (8) furnishes a symmetrical displacement field $\underline{\Gamma}^l_{ik}$ and a skew-symmetrical tensor $\overset{\vee}{\Gamma}{}^l_{ik}$ whose additive composition appears at first unwarranted. However, we shall presently see that this objection, just as in the analogous case of the non-symmetrical tensor g_{ik}, has no decisive significance.

If we define

$$\tilde{\Gamma}^l_{ik} = \Gamma^l_{ki} \tag{9}$$

it follows from (9) that

$$\tilde{\Gamma}^l_{ik} = \underline{\Gamma}^l_{ik} + (-\overset{\vee}{\Gamma}{}^l_{ik})$$

Because of the transformation formula (7), we see that $\tilde{\Gamma}^l_{ik}$ is likewise a displacement field. In an analogous way, one obtains, of course, the tensor $\tilde{g}_{ik} = g_{ki}$ by transposition of the g_{ik}. The simultaneous discussion of g, Γ and their transposed $\tilde{g}$, $\tilde{\Gamma}$ will prove important in the sequel.

Since the displacement $\tilde{\Gamma}$ is given together with a displacement Γ the expressions (6a), (6b) involve two different kinds of absolute derivatives which must be kept apart. Hence, we define

$$\left.\begin{aligned} A^i_{+;k} &= A^i{}_{,k} + A^s\Gamma^i_{sk} & A_{i;k}_{+} &= A_{i,k} - A_s\Gamma^s_{ik} \\ A^i_{-;k} &= A^i{}_{,k} + A^s\Gamma^i_{ks} & A_{i;k}_{-} &= A_{i,k} - A_s\Gamma^s_{ki} \end{aligned}\right\} \tag{10a, 10b}$$

Furthermore, it happens occasionally that the absolute derivatives appear with symmetrized Γ. Hence, we define

$$A^i_{\cdot;k} = A^i{}_{,k} + A^s\underline{\Gamma}^i_{sk} \qquad A_{i;k}_{\circ} = A_{i,k} - A_s\underline{\Gamma}^s_{ik} \tag{11a, 11b}$$

As in the case of symmetrical Γ, the sum and product rules hold for absolute differentiation and imply, in particular, the rules for the differentiation of tensors of higher rank.

One should take note of the following circumstance for which no analog exists in the calculus of symmetrical Γ: the absolute

APPENDIX II

derivative of the Kronecker tensor δ_i^k vanishes only in the cases

$$\underset{+}{\delta_i^k}_{;l} \quad , \quad \underset{-}{\delta_i^k}_{;l} \quad , \quad \underset{\circ}{\delta_i^k}_{;l}$$

while, for example,

$$\underset{+}{\delta_i^k}_{-;l} = 0 + \delta_i^s \Gamma_{ls}^k - \delta_s^k \Gamma_{il}^s = 2\underset{\circ}{\Gamma_{li}^k} \neq 0.$$

For this reason, contraction under the sign of differentiation (without adding correction terms) is permissible *only* if the respective indices have the same character with respect to differentiation. Thus, for example,

$$\underset{+}{A^i}_{k\ldots;l}\, \delta_i^k = A_{s\ldots;l}^s .$$

Absolute Differentiation of Tensor Densities.

$$\underset{+-}{g_{ik;l}} \; (= g_{ik,l} - g_{sk}\Gamma_{il}^s - g_{is}\Gamma_{lk}^s) \tag{12}$$

is a tensor. By multiplication with g^{ik} and contraction with respect to i and k we obtain, according to the rule of differentiation for determinants, the vector

$$\frac{w_{,l}}{w} - \underline{\Gamma_{ls}^s} \tag{13}$$

where w designates $\sqrt{-\mathrm{Det}(g_{ik})}$. w is a scalar density (of degree 1). Furthermore, if ρ be any scalar density (of degree 1), then $\frac{\rho}{w}$, hence also $\log \frac{\rho}{w}$, is a scalar. By differentiation with respect to x_l the vector

$$\underline{\frac{\rho_{,l}}{\rho}} - \frac{w_{,l}}{w} \tag{14}$$

is obtained. Adding the vectors (13) and (14) we obtain the vector

$$\underline{\frac{\rho_{,l}}{\rho}} - \underline{\Gamma_{ls}^s} .$$

APPENDIX II

Multiplication by ρ yields the vector density

$$\rho_{,l} - \rho\Gamma^{s}_{\underline{ls}} .$$

This we define as the absolute derivative $\rho_{;l}$ of the scalar density ρ:

$$\rho_{;l} = \rho_{,l} - \rho\Gamma^{s}_{\underline{ls}}. \tag{15}$$

From this definition it follows, in particular, that the scalar density w satisfies the identity

$$\left(\frac{w_{;\underline{l}}}{w}\right)_{,m} - \left(\frac{w_{;\underline{m}}}{w}\right)_{,l} \equiv \Gamma^{s}_{\underline{ls},m} - \Gamma^{s}_{\underline{ms},l} \tag{16}$$

Using (15) and applying the product rule for absolute derivatives we may extend absolute differentiation in the familiar way to tensor densities. Thus, for example,

$$(\rho A^{i}_{+})_{;k} = \rho_{;k}A^{i} + \rho A^{i}_{\ ;k}$$

or, denoting the vector density ρA^{i} by $\mathfrak{A}^{i}$,

$$\mathfrak{A}^{i}_{+;k} = (\rho_{,k} - \rho\Gamma^{s}_{\underline{ks}})A^{i} + \rho(A^{i}_{\ ,k} + A^{s}\Gamma^{i}_{sk})$$

or

$$\mathfrak{A}^{i}_{+;k} = \mathfrak{A}^{i}_{\ ,k} + \mathfrak{A}^{s}\Gamma^{i}_{sk} - \mathfrak{A}^{i}\Gamma^{s}_{\underline{ks}}. \tag{17}$$

The last term distinguishes the absolute derivative of a vector density from that of a vector. An analogous result holds for the absolute differentiation of tensor densities of higher rank. By contracting (17) one obtains

$$\mathfrak{A}^{i}_{+;i} = \mathfrak{A}^{i}_{\ ,i} + \mathfrak{A}^{i}\Gamma_{i} \tag{18}$$

where we introduce $\Gamma_{i} = \Gamma^{s}_{\underline{is}}$. Correspondingly we obtain

$$\mathfrak{A}^{i}_{-;i} = \mathfrak{A}^{i}_{\ ,i} - \mathfrak{A}^{i}\Gamma_{i} \tag{19}$$

Identities of the Metric Tensor

We shall now derive several identities which will be useful in what follows. As in the previous section, we multiply the

APPENDIX II

tensor $\underset{+-}{g_{ik;l}}$ by g^{ik} and contract. Using the definition of $w_{;l}$ we can now write:

$$g^{ik}\underset{+-}{g_{ik;l}} \equiv 2w_{;l} \tag{20}$$

Again, we multiply $\underset{+-}{g_{ik;l}}$ by $g^{it}g^{sk}$ and contract. This gives (using $g^{sk}g_{ik,l} = -g^{sk}{}_{,l}g_{ik}$):

$$g^{it}g^{sk}\underset{+-}{g_{ik;l}} \equiv -g^{st}{}_{,l} - g^{it}\Gamma^{s}_{il} - g^{sk}\Gamma^{t}_{lk} \equiv -\overset{s\ t}{\underset{+-}{g}}{}_{;l} \tag{21}$$

Finally, we define $\mathfrak{g}^{ik} = w\cdot g^{ik}$. For this tensor density we have

$$\overset{i\ k}{\underset{+-}{\mathfrak{g}}}{}_{;l} = \mathfrak{g}^{ik}{}_{,l} + \mathfrak{g}^{sk}\Gamma^{i}_{sl} + \mathfrak{g}^{is}\Gamma^{k}_{ls} - \mathfrak{g}^{ik}\Gamma^{s}_{ls}$$

Anti-symmetrizing with respect to i and k, and contracting with respect to k and l, we get the important identity:

$$\tfrac{1}{2}(\mathfrak{g}^{it}{}_{;t} - \mathfrak{g}^{ti}{}_{;t}) \equiv \mathfrak{g}^{it}{}_{,t} - \mathfrak{g}^{it}\Gamma_{t} \tag{22}$$

Riemann Curvature

Just as in the case of the symmetrical theory, a curvature tensor may be obtained by parallel displacement of a vector along the boundary of an infinitesimal surface element:

$$R^{i}{}_{klm} = (\Gamma^{i}_{kl,m} - \Gamma^{i}_{sl}\Gamma^{s}_{km}) - (\Gamma^{i}_{km,l} - \Gamma^{i}_{sm}\Gamma^{s}_{kl}) \tag{23}$$

Another tensor of this kind would be obtained by replacing Γ^{i}_{kl} by $\bar{\Gamma}^{i}_{kl}$ $(=\Gamma^{i}_{lk})$.

Contraction of (23) with respect to i and m results in the tensor

$$R_{kl} = (\Gamma^{t}_{kl,t} - \Gamma^{t}_{sl}\Gamma^{s}_{kt}) - (\Gamma^{t}_{kt,l} - \Gamma^{t}_{st}\Gamma^{s}_{kl}) \tag{24}$$

§3. DERIVATION OF THE FIELD EQUATIONS

In order to get a system of equations which is compatible, we restrict ourselves to the case in which the system of equations is derivable from a variational principle. This guarantees the compatibility of the equations.

APPENDIX II

We choose as the most general integrand for our variational principle a scalar density $\mathfrak{H}$, built out of the $\mathfrak{g}^{ik}$, Γ^{l}_{ik} and their first derivatives. The variational principle then says:

$$\delta \int \mathfrak{H}\, d\tau = 0 \tag{25}$$

where $\mathfrak{g}^{ik}$ and Γ^{l}_{ik} are to be varied independently of one another. We can choose $\delta\mathfrak{g}$ and $\delta\Gamma$ so that they vanish at the boundaries of integration. Then, after partial integration, we can rewrite (25) in the form

$$\int (\mathfrak{B}^{ik}{}_{l}\,\delta\Gamma^{l}_{ik} + W_{ik}\cdot\delta\mathfrak{g}^{ik})\, d\tau = 0 \tag{25a}$$

Since the $\delta\mathfrak{g}$ and $\delta\Gamma$ are independent, their coefficients must vanish, and we get the field-equations

$$\left.\begin{aligned} \mathfrak{B}^{ik}{}_{l} &= 0 \\ W_{ik} &= 0 \end{aligned}\right\} \qquad \text{(I)}$$

Now consider the integral of (25) to be varied under the a priori condition

$$\mathfrak{g}^{is}{}_{,s} = 0 \tag{26}$$

This relationship among the $\mathfrak{g}^{ik}$, which is given a priori, means that the $\delta\mathfrak{g}^{ik}$ are no longer independent. In order to take this into account we can use, for example, the method of Lagrange multipliers. We then find, after partial integration:

$$\int (\mathfrak{B}^{ik}{}_{l}\,\delta\Gamma^{l}_{ik} + [W_{ik} + \sigma_{i,k} - \sigma_{k,i}]\delta\mathfrak{g}^{ik})\, d\tau = 0$$

where σ_i is the vector field of the Lagrange multipliers, which enables us now to treat the $\delta\mathfrak{g}^{ik}$ as independent variations. We are now led to the equations

$$\mathfrak{B}^{ik}{}_{l} = 0, \qquad \mathfrak{g}^{is}{}_{,s} = 0, \qquad W_{ik} + (\sigma_{i,k} - \sigma_{k,i}) = 0$$

APPENDIX II

or, eliminating the auxiliary σ_i :

$$\left.\begin{aligned} &\mathfrak{B}^{ik}{}_{l} = 0, \qquad \mathfrak{g}^{is}_{\ \ ,s} = 0 \\ &W_{\underset{\smile}{ik}} = 0, \qquad W_{\underset{\smile}{ik},l} + W_{\underset{\smile}{kl},i} + W_{\underset{\smile}{li},k} = 0 \end{aligned}\right\} \qquad \text{(II)}$$

The system of equations (II) differs from (I) in that the six equations $W_{\underset{\smile}{ik}} = 0$ of (I) are replaced by eight equations:

$$\mathfrak{g}^{is}_{\ \ ,s} = 0, \qquad W_{\underset{\smile}{ik},l} + W_{\underset{\smile}{kl},i} + W_{\underset{\smile}{li},k} = 0$$

Among these eight equations, two identities hold:

$$(\mathfrak{g}^{is}_{\ \ ,s})_{,i} \equiv 0$$

and

$$(W_{\underset{\smile}{ik},l} + W_{\underset{\smile}{kl},i} + W_{\underset{\smile}{li},k})_{,m} \cdot \eta^{iklm} \equiv 0$$

(η^{iklm} is the Levi-Civita tensor-density, which is anti-symmetric in all four indices).

It seems evident (and this will be proved rigorously below, see p. 160), that these eight equations with two identities give a stronger determination of the field variables than the six equations of (I), for whatever tensor-density $\mathfrak{H}$ we choose in our variational principle. I feel compelled to prefer the "stronger" system (II), as a physical theory, over system (I). This principle, of preferring stronger systems to weaker ones, seems to me to be of general importance.

"Bianchi" Identities

In the purely gravitational theory we know the four Bianchi identities. These can be obtained, following H. Weyl, from a variational principle, making use of the invariance with respect to the group of coordinate transformations. The same procedure can be followed in the non-symmetrical theory. We suppose the variations $\delta\Gamma^{l}_{ik}$ and $\delta\mathfrak{g}^{ik}$ to arise from an infinitesimal change of coordinates. Then we have*

$$\delta\Gamma^{l}_{ik} = \Gamma^{s}_{ik}\xi^{l}_{\ ,s} - \Gamma^{l}_{sk}\xi^{s}_{\ ,i} - \Gamma^{l}_{is}\xi^{s}_{\ ,k} - \xi^{l}_{\ ,ik} - \Gamma^{l}_{ik,s}\xi^{s}$$
$$\delta\mathfrak{g}^{ik} = \mathfrak{g}^{sk}\xi^{i}_{\ ,s} + \mathfrak{g}^{is}\xi^{k}_{\ ,s} + \mathfrak{g}^{ik}\xi^{s}_{\ ,s} - \mathfrak{g}^{ik}_{\ \ ,s}\xi^{s}$$

* The last term in each line comes from the fact that the varied quantities have to be taken for fixed values of the coordinates.

APPENDIX II

Putting these into the varied integral $\delta\int \mathfrak{H}\, d\tau = 0$, and carrying out the necessary partial integrations, we can bring the integral into the form

$$\int \mathfrak{W}_i \xi^i \, d\tau = 0 \tag{27}$$

hence we have four "Bianchi" identities:

$$\mathfrak{W}_i \equiv 0 \tag{28}$$

The explicit content of these identities depends of course on the choice of $\mathfrak{H}$ in the variational principle. However, even before making this choice, we can see how the identities in (28) connect the field equations arising from the general variational principle $\delta\int \mathfrak{H}\, d\tau = 0$. Indeed, when we put $\delta\Gamma^l_{ik}$ and $\delta\mathfrak{g}^{ik}$ into (25a), and bring into the form (27), we see that $\mathfrak{W}_i$ contains $\mathfrak{B}^{ik}{}_l$ and W_{ik} linearly, together with their first derivatives, so that if the system (I) of field equations is satisfied, (28) are identities among these field equations. In the case of system (II) the same is true, but to see this one has to separate out the symmetric and antisymmetric parts of W_{ik} which appear in $\mathfrak{W}_i$. The contribution of W_{ik} to $\mathfrak{W}_i$ is

$$[-2(W_{\underline{si}}\mathfrak{g}^{\underline{st}})_{,t} + W_{\underline{sk}}{}_{,i}\mathfrak{g}^{\underline{sk}}] + \mathfrak{g}^{sk}\cdot[W_{\underset{\smile}{k\,i,s}} + W_{\underset{\smile}{is,k}} + W_{\underset{\smile}{sk,i}}] + [-W_{\underset{\smile}{ik}}\mathfrak{g}^{\underset{\smile}{tk}}{}_{,t} + W_{\underset{\smile}{ik}}\mathfrak{g}^{\underset{\smile}{kt}}{}_{,t}]$$

The last bracket is just $2W_{\underset{\smile}{sk}}\cdot\mathfrak{g}^{\underset{\smile}{kt}}{}_{,t}$, ($=0$ because of (26)). So we see that $\mathfrak{W}_i$ is a linear combination of the field equations also in the case of system (II).

Choice of Hamiltonian $\mathfrak{H}$

We are left with the question of how to choose the tensor density $\mathfrak{H}$ in our variational principle. Nothing has been said about $\mathfrak{H}$ up to now. In order to enable us to make a choice, we impose the following condition on the variational principle: it

APPENDIX II

should lead (if possible) to the simplest relationship between g and Γ, namely:

$$g_{\underset{+-}{ik;l}} = 0. \tag{29}$$

(29) is a generalization of the relationship holding in the symmetrical theory. It has the property of transposition invariance (which is of course meaningless in the symmetrical theory); that is, the equation still holds when we replace g_{ik} by $\tilde{g}_{ik}(\equiv g_{ki})$, and Γ^i_{kl} by $\tilde{\Gamma}^i_{kl}$ $(\equiv \Gamma^i_{lk})$. This fact is the mathematical expression of the indifference of the theory to the sign of electricity.

A density $\mathfrak{H}$ fulfilling the condition stated above can indeed be found. We start from the contracted curvature tensor:

$$R_{ik} = [\Gamma^s_{ik,s} - \Gamma^t_{is}\Gamma^s_{tk}] - [\Gamma^s_{is,k} - \Gamma^t_{ik}\Gamma^s_{ts}] \tag{24}$$

To R_{ik} we will first add the tensor $\Gamma_{\underset{+}{i;k}} = \Gamma^s_{\underset{\smile}{is},k} - \Gamma^t_{ik}\Gamma^s_{\underset{\smile}{ts}}$; this replaces R_{ik} by

$$R_{ik}{}^* = [\Gamma^s_{ik,s} - \Gamma^t_{is}\Gamma^s_{tk}] - [\Gamma^s_{\underline{is}},_k - \Gamma^t_{ik}\Gamma^s_{\underline{ts}}]$$

In the second bracket we now make use of the equation defining absolute differentiation of a scalar density w:

$$w_{;t} = w_{,t} - w \cdot \Gamma^s_{\underline{ts}} \tag{15}$$

This gives, after rearranging terms,

$$R^*_{ik} = [\Gamma^s_{ik,s} - \Gamma^t_{is}\Gamma^s_{tk}] - \left[\left(\frac{w_{,i}}{w}\right)_{,k} - \left(\frac{w_{,t}}{w}\right)\Gamma^t_{ik}\right] + \left[\left(\frac{w_{;\,i}}{w}\right)_{,k} - \left(\frac{w_{;\,t}}{w}\right)\Gamma^t_{ik}\right].$$

The last bracket is simply the tensor $\left(\frac{1}{w}\cdot w_{\underset{+}{;\,i}}\right)_{;k}$. Subtracting this tensor from R^*_{ik} we get

$$R^{**}_{ik} = \Gamma^s_{ik,s} - \Gamma^t_{is}\Gamma^s_{tk} - (\log w)_{,ik} + (\log w)_{,t}\Gamma^t_{ik} \tag{24a}$$

By its construction, R^{**}_{ik} is a tensor; and indeed, we shall now

APPENDIX II

verify that the density $\mathfrak{g}^{ik}R^{**}_{ik}$ can serve for our variational principle, and lead to the desired result. We therefore vary the integral $\int \mathfrak{g}^{ik}R^{**}_{ik}\, d\tau$, first with respect to Γ (where the auxiliary condition $\mathfrak{g}^{is}_{\ \ ,s} = 0$ plays no role). We get the equation

$$\mathfrak{g}^{ik}_{\ \ ,l} + \mathfrak{g}^{tk}\Gamma^{i}_{tl} + \mathfrak{g}^{it}\Gamma^{k}_{lt} - \mathfrak{g}^{ik}(\log w)_{,l} = 0 \tag{30}$$

Replacing $\mathfrak{g}^{ik}$ by $w \cdot g^{ik}$ we find, after dividing throughout by w:

$$0 = g^{ik}_{\ \ ,l} + g^{tk}\Gamma^{i}_{tl} + g^{it}\Gamma^{k}_{lt} = \overset{i\,k}{\underset{+-}{g}}_{;l} \tag{30a}$$

which, according to identities (20) and (21), results in

$$g_{\underset{+-}{ik};l} = 0\,, \qquad \overset{i\,k}{\underset{+-}{\mathfrak{g}}}_{;l} = 0\,, \qquad w_{;l} = 0$$

as desired.

Now that we have the two equations $\overset{i\,k}{\underset{+-}{\mathfrak{g}}}_{;l} = 0$ and $g^{is}_{\ \ ,s} = 0$, we see from the identity (22), that also the equations

$$\Gamma_i = 0 \tag{31}$$

are fulfilled. And this has the further consequence that our R^{**}_{ik} is now equal to R_{ik}, since the difference between the two tensors is the vanishing quantity

$$\Gamma_{\underset{+}{i\,k}} - \left(\frac{1}{w}\cdot w_{;\underset{+}{i}}\right)_{;k}$$

Finally, we get the remaining equations by varying $\int \mathfrak{g}^{ik}R^{**}_{ik}\, d\tau$ with respect to $\mathfrak{g}^{ik}$. We have, after partial integration,

$$0 = \int \left[R^{**}_{ik}\,\delta\mathfrak{g}^{ik} - \{\mathfrak{g}^{ik}_{\ \ ,ik} + (\mathfrak{g}^{ik}\Gamma^{t}_{ik})_{,t}\}\delta(\log w)\right] d\tau$$

But the coefficient of $\delta(\log w)$ vanishes, for we can rewrite it as

$$\{\mathfrak{g}^{ik}_{\ \ ,i} + \mathfrak{g}^{it}\Gamma^{k}_{it}\}_{,k} = \{\mathfrak{g}^{ik}_{\ \ ;i}\}_{,k} = 0;$$

APPENDIX II

hence we are left with

$$\int R_{ik}^{**}\,\delta\mathfrak{g}^{ik}\,d\tau = 0$$

The $\delta\mathfrak{g}^{ik}$ do not vary independently, because of the a priori condition $\mathfrak{g}^{is}_{\ \ ,s} = 0$, and we use Lagrange multipliers, as shown above. The resulting equations are (we can now write R_{ik} in place of R_{ik}^{**}):

$$R_{ik} = 0, \qquad R_{ik,l} + R_{kl,i} + R_{li,k} = 0.$$

In writing down the complete system of equations, we note that equations $\mathfrak{g}^{ik}_{+-;l} = 0$ with $\Gamma_i = 0$ are together equivalent to $\mathfrak{g}^{ik}_{+-;l} = 0$ with $\mathfrak{g}^{is}_{\ \ ,s} = 0$, because of identity (22). Hence we can choose for our system:

$$\left.\begin{aligned} \mathfrak{g}^{ik}_{+-;l} &= 0, \qquad \Gamma_i = 0 \\ R_{ik} &= 0, \qquad R_{ik,l} + R_{kl,i} + R_{li,k} = 0 \end{aligned}\right\} \qquad \text{(IIa)}$$

This system of equations is transposition invariant; this is seen immediately for the equations $\mathfrak{g}^{ik}_{+-;l} = 0$, $\Gamma_i = 0$. For the other two, the easy way to check on the transposition invariance is to substitute R_{ik}^{**} for R_{ik}.

We have now to ask whether some other $\mathfrak{H}$ could have been chosen which would also have led to the equation $g^{ik}_{+-;l} = 0$. Suppose we had found another such density $\mathfrak{H}^*$, such that $\mathfrak{H}^* = \mathfrak{H} + \mathfrak{K}$. If $\mathfrak{K}$ depends on the Γ, it will contribute to the variation with respect to Γ. On the other hand, this contribution must vanish identically, in order to preserve the equation $g^{ik}_{+-;l} = 0$. We may trust without proof that this means that $\mathfrak{K}$ cannot depend on Γ. The possibility remains that $\mathfrak{K}$ is built out of g_{ik} alone, for example in the forms w, or $wg^{ik}g_{ki}$, etc. (The first of these possibilities leads to the so-called "cosmological term.") All such additional terms bring a heterogeneity into the system of equations, and can be disregarded, provided that no strong physical argument is found to support them.

APPENDIX II

Remarks about the physical interpretation of the system of equations. The question which comes to the foreground concerning the physical interpretation of the theory is: To what extent is there an affinity between this theory and Maxwell's? There exists a formal similarity between the linear approximation to this theory and the Maxwell equations. One finds, in the linear approximation, that the system decomposes into two sets of equations, one set for the symmetric components of the field, and the other for the antisymmetric components. The equations for the antisymmetric part constitute a generalization of Maxwell's equations for empty space. This decomposition, which occurs in the linear approximation, makes it understandable why the electromagnetic and gravitational fields seemed so independent of one another in the previous development of our theoretical ideas. For those ideas resulted exclusively from our knowledge about the behavior of weak fields.

In the rigorous theory this independence no longer holds. The relationship to Maxwell's theory is in any case not very close; in particular, the localization of the energy is quite different. However, the following analogies present themselves:

The equation $\mathfrak{g}^{is}_{\ \ ,s} = 0$ suggests that it can be interpreted as the expression for the vanishing of the magnetic current. Furthermore, it is plausible to interpret the quantities

$$g_{ik,l} + g_{kl,i} + g_{li,k}$$

or the corresponding vector-density

$$\mathfrak{J}^m = \tfrac{1}{6}(g_{ik,l} + g_{kl,i} + g_{li,k})\eta^{iklm}$$

as the electric current density, since these quantities satisfy the "continuity equation"

$$\mathfrak{J}^m{}_{,m} = 0$$

identically.

An Alternative Derivation of the Field Equations

The derivation given above for the field equations, which makes use of the auxiliary tensor R^{**}_{ik}, has several advantages.

APPENDIX II

The variation with respect to Γ gives the equation $g_{\underset{+-}{ik;l}} = 0$ directly. Furthermore, the transposition invariance is immediately seen, when we write R^{**}_{ik} for R_{ik} in the equations.

On the other hand, however, there are disadvantages to this derivation: R^{**}_{ik} is introduced in an artificial way. Also, from the derivation it would seem that an equally justified alternative is to choose the antisymmetrical parts of the field (g_{ik} and Γ^l_{ik}) as purely imaginary. It may not be superfluous, therefore, to sketch another method of derivation which avoids these shortcomings.

We assign to each Γ^l_{ik} field a family of fields $\Gamma^{l\dagger}_{ik}$ which are connected with one another through the relation

$$\Gamma^{l\dagger}_{ik} = \Gamma^l_{ik} + \delta^l_i \lambda_k \qquad (\text{"}\lambda\text{-transformation"})$$

where λ_k is an arbitrary vector-field.* It is easy to show that the coordinate- and λ-transformations together build a group (the "extended group"). The question is now, whether one can find a variational principle (again with the a priori condition $\mathfrak{g}^{is}_{\underset{\vee}{\ },s} = 0$), which is invariant under the extended group.

When we apply the λ-transformation to R_{ik}, we find

$$R^{\dagger}_{ik} = R_{ik} - (\lambda_{i,k} - \lambda_{k,i})$$

Hence, the tensor-density $\mathfrak{H} = \mathfrak{g}^{ik} R_{ik}$ is transformed in this way:

$$(\mathfrak{g}^{ik} R^{\dagger}_{ik} =)\mathfrak{H}^{\dagger} = \mathfrak{H} - \mathfrak{g}^{ik}(\lambda_{i,k} - \lambda_{k,i}) = \mathfrak{H} + 2\mathfrak{g}^{\underset{\vee}{ik}}\lambda_{k,i}$$

Because of the condition $0 = \mathfrak{g}^{is}_{\underset{\vee}{\ },s}$, we can rewrite this as

$$\mathfrak{H}^{\dagger} = \mathfrak{H} + 2(\mathfrak{g}^{\underset{\vee}{ik}}\lambda_k)_{,i}$$

The second term, when integrated, gives a surface integral, and contributes therefore nothing to the variation. Hence the variational principle

$$\delta\left\{\int \mathfrak{H}\, d\tau\right\} = 0$$

* Clearly this transformation "mixes" the symmetrical and antisymmetrical parts of the Γ-field, so that one cannot choose $\Gamma^i_{\underset{\vee}{kl}}$ to be imaginary here.

APPENDIX II

is invariant under the λ-transformation, as are also the field equations resulting from it.

The variation with respect to Γ can be shown to lead to

$$-\mathfrak{g}^{\overset{ik}{+-}}{}_{;l} + \mathfrak{g}^{\overset{it}{+}}{}_{;t}\delta^k_l + \mathfrak{g}^{ik}\Gamma_l + \mathfrak{g}^{it}\Gamma_t\delta^k_l = 0$$

From this it follows that $\mathfrak{g}^{\overset{ik}{+-}}{}_{;l} = 0$ if $\Gamma_i = 0$ is fulfilled. But this can (because of the λ-invariance of the variational-principle) be achieved by a suitable choice of λ_k ("normalization" of the Γ-field).

We have thus arrived at the same equations as in the previous method. Here, however, it is essential that Γ^l_{ik} be real. Also $\Gamma^{\overset{\vee}{l}}_{ik}$ has no longer an invariant meaning under the extended group. When we require here again that the equation $g_{\overset{ik}{+-};l} = 0$ hold, then, by the same arguments as given above (p. 146) we see that the Hamiltonian function can have no further terms added to it.

§4. GENERAL REMARKS ABOUT THE "STRENGTH" OF EQUATION-SYSTEMS. APPLICATION TO THE THEORY OF THE NON-SYMMETRICAL FIELD

In the previous section (p. 142), we gave preference to the system of equations (II) over system (I). The principle we followed there is applicable in general to physical theories: The system of equations is to be chosen so that the field quantities are determined *as strongly as possible*.

In order to apply this principle, we need a method which gives a measure of the strength of an equation system. We do as follows: We expand the field variables, in the neighborhood of a point P, into Taylor series (which presupposes the analytic character of the field). The coefficients of these series, which are simply the derivatives of the field variables at the point P, fall into sets according to the degree of differentiation. In every such degree there appear, for the first time, a set of coefficients which would be free for arbitrary choice if it were not that the

APPENDIX II

field must satisfy a system of differential equations. Through this system of differential equations (and its derivatives with respect to the coordinates) the number of coefficients is restricted, so that in each degree a smaller number of coefficients is left free for arbitrary choice. The set of numbers of "free" coefficients for all degrees of differentiation is then directly a measure of the "weakness" of the system of equations, and through this, indirectly, also of its "strength."

This method of counting the free coefficients in each degree, presupposes that these numbers can be successively calculated degree after degree. We can also restrict ourselves to the case that all these numbers come out to be positive, that is to say that the equations for the coefficients of one degree do not work backwards to require, for their fulfillment, more conditions on the coefficients of the previous degrees. In such a case we refer to the system of equations as "absolutely compatible." The systems of differential equations used hitherto in theoretical physics seem all to have this character.

We will now illustrate this method by several examples. The wave equation in four dimensions:

$$0 = \Box\phi = \phi_{,11} + \phi_{,22} + \phi_{,33} - \phi_{,41}$$

The system of equations consists here of *one* equation for *one* function ϕ. One has, then, for ϕ, in the neighborhood of a point, say $x_i = 0$, a power series of the form

$$\phi = c + (c_i x_i) + (c_{ik} x_i x_k) + \cdots$$

ϕ is then determined through the statement of the values for the coefficients c in all degrees of the expansion (i.e. through the derivatives of ϕ at $x_i = 0$). The number of coefficients in each successive degree in the calculation is given in the table below:

Degree	Number of coefficients
0	1
1	$4 = \binom{4}{1}$

APPENDIX II

$$\begin{array}{ll} 2 & \dfrac{4\cdot 5}{1\cdot 2} = \binom{4}{2} \\ 3 & \dfrac{4\cdot 5\cdot 6}{1\cdot 2\cdot 3} = \binom{4}{3} \\ \cdots & \cdots\cdots \\ n & \dfrac{4\cdot 5\cdots(n+3)}{1\cdot 2\cdots n} = \binom{4}{n}. \end{array}$$

The number of these coefficients is diminished through the conditions which follow from the differential equation. These conditions are obtained by taking into account the differential equation and all its derivatives at the point $x_i = 0$.

Degree	Number of conditions
0	0
1	0
2	1
3	$\binom{4}{1}$
...	.....
n	$\binom{4}{n-2}$

When we subtract, for each degree, the number of conditions from the number of coefficients, we obtain the number of coefficients, Ω_n, in each degree which remain free for arbitrary choice. For the degree n (>2) we find the number of such free coefficients to be

$$\Omega_n = \binom{4}{n} - \binom{4}{n-2}$$

or

$$\Omega_n = \binom{4}{n}\left(1 - \frac{n}{n+3}\cdot\frac{n-1}{n+2}\right) = \binom{4}{n}\left(1 - \frac{1}{1+\frac{3}{n}}\cdot\frac{1-\frac{1}{n}}{1+\frac{2}{n}}\right)$$

For large n, this can be replaced by

$$\Omega_n \sim \binom{4}{n}\cdot\frac{6}{n}$$

APPENDIX II

The factor $\frac{6}{n}$ gives the fraction of the number of coefficients (for the degree $n \gg 1$), which remain undetermined by the differential equation. Ω_n is then a measure of the strength of the determination of the field, through the differential equation. Ω_n, being a positive number, characterizes this system as an "absolutely compatible" one.

These considerations can be carried over without any change to the case of the more complicated equation

$$A_{ik} \cdot \phi_{,ik} + B = 0$$

where A_{ik} and B are given functions of ϕ and $\phi_{,s}$. When we differentiate this equation to the degrees $(n - 2)$, at the point $x_i = 0$, we find that the nth derivatives of ϕ appear linearly, while the lower derivatives of ϕ enter in a more complicated form. In the lower degrees of differentiation of the equation, the nth derivative of ϕ does not appear at all. This is what makes it possible for the various steps of differentiation to be fulfilled successively. The schema showing the number of conditions which follow from the differential equation, gives, in the right-hand column, the number of equations in which the degree of differentiation, listed in the lefthand column, appears for the first time (and appears linearly).

The main complication which arises in the application of the method sketched above to systems of more than one differential equations consists of taking into account correctly identities among the equations. We show how to do this with the example of Maxwell's equations for the vacuum.

Maxwell's Equations for Empty Space

We write these in Minkowski's form (with imaginary time-coordinate):

$$(U_i \equiv) \phi_{is,s} = 0$$

$$(V_{ikl} \equiv) \phi_{ik,l} + \phi_{kl,i} + \phi_{li,k} = 0$$

APPENDIX II

where ϕ_{ik} is the (antisymmetric) tensor of the electromagnetic field. These equations satisfy the identities

$$U_{\iota,i} = 0$$

$$V_{ikl,m}\eta^{iklm} = 0$$

ϕ_{ik} has six components. $6\binom{4}{n}$ is therefore the number of Taylor-series coefficients in the nth step. Connecting these coefficients are algebraic conditions, which come from the $(n-1)$th differentiation of the field equations. The number of these conditions is

$$4\binom{4}{n-1} + 4\binom{4}{n-1}$$

However, because of the existence of identities, these conditions are not independent. Indeed, there are relations among the coefficients of the nth step which hold identically in consequence of the $(n-2)$ differentiation of the identities listed above. The number of these identical relations among the coefficients is

$$\binom{4}{n-2} + \binom{4}{n-2}$$

Hence, the number of *independent* relations among the coefficients of the nth step is

$$8\binom{4}{n-1} - 2\binom{4}{n-2}$$

When we subtract this number from the (total) number, $6\cdot\binom{4}{n}$, of coefficients in the nth step, we get the number Ω_n of coefficients which remain undetermined in the nth step:

$$\Omega_n = 6\binom{4}{n} - \left[8\binom{4}{n-1} - 2\binom{4}{n-2}\right]$$

APPENDIX II

It is convenient at this point to use the asymptotic value ($n \gg 1$)

$$\binom{4}{n-r} \sim \binom{4}{n}\left(1 - \frac{3r}{n}\right) \qquad (r \ll n)$$

It follows now that

$$\Omega_n \sim \binom{4}{n}\left[6 - 8\left(1 - \frac{3}{n}\right) + 2\left(1 - \frac{6}{n}\right)\right]$$

or

$$\Omega_n \sim \binom{4}{n} \cdot \frac{12}{n}$$

We see that there remain here, in the nth step, twice as many free coefficients as in the case of the scalar wave-equation.

Before we leave this example, we want to mention one point, which is needed both for the understanding of the method used here and for the formal understanding of this particular system of equations.

When we introduce, into the equations above, the vector-potential ψ_i, according to the definition

$$\phi_{ik} = \psi_{i,k} - \psi_{k,i}$$

we find that the second equation is automatically fulfilled, while the first gives

$$(W_i \equiv)\psi_{i,s,s} - \psi_{s,s,i} = 0$$

These are four equations for the four components ψ_i, and among these equations the identity $W_{i,i} \equiv 0$ holds. This is, thus, an incomplete determination for the vector-field ψ_i, out of which we get, conversely, the original system of equations. A "complete" determination of the ψ_i-field is obtained, as is known, in adjoining to this (intermediary) system the equation $\psi_{s,s} = 0$. One gets then the system

$$(A_i \equiv)\psi_{i,s,s} = 0$$
$$(B \equiv)\psi_{s,s} = 0$$

APPENDIX II

The identity given above then takes the form

$$A_{i,i} - B_{,s,s} \equiv 0$$

This system implies the intermediary system, and with it also the original equations for the ϕ_{ik} . Hence this system determines the ϕ_{ik} with the same strength as does the original system; it is however also of interest to see with what strength ψ_i are determined. The counting method for the coefficients of the nth degree gives here:

$$\Omega_n = 4\binom{4}{n} - \left[4\binom{4}{n-2} + \binom{4}{n-1} - \binom{4}{n-3}\right]$$

or, asymptotically,

$$\Omega_n \sim \binom{4}{n}\left(\frac{24}{n} + \frac{3}{n} - \frac{9}{n}\right) = \binom{4}{n}\frac{18}{n}$$

A comparison with the results from the original system of equations shows that the ψ_i-field is determined more weakly than the ϕ_{ik}-field. In this sense the original form of the Maxwell equations is to be preferred, if there is no reason to give an independent physical meaning to the variables ψ_i . The reason why we find different strengths of determination in these two systems of equations is that the ψ_i , even with completely known ϕ_{ik} , are not fully determined. We will show this in closer detail, since this case enables us to demonstrate how to take into consideration "identities of identities" during the counting process. Let ϕ_{ik} be given, and the ψ_i satisfy the equations

$$(C_{ik} \equiv) \psi_{i,k} - \psi_{k,i} = \phi_{ik} \qquad \text{(where the given } \phi_{ik} \text{ satisfy}$$

$$(B \equiv) \psi_{s,s} = 0 \qquad \phi_{ik,l} + \phi_{kl,i} + \phi_{li,k} = 0)$$

The first of these equations satisfies the identity

$$(D_{ikl} \equiv) C_{ik,l} + C_{kl,i} + C_{li,k} \equiv 0$$

These four identities, however, are not independent of one another, since the D_{ikl} themselves fulfill an identity, for what-

APPENDIX II

ever quantities the (antisymmetric) C_{ik} may be; namely, the identity

$$D_{ikl,m}\eta^{iklm} \equiv 0$$

This identity of the D_{ikl} changes the result of counting of coefficients in every step. Without giving the proof here, we state the result:

$$\Omega_n = 4\binom{4}{n} - \left[\left\{6\binom{4}{n-1} + \binom{4}{n-1}\right\} - \left\{4\binom{4}{n-2} - \binom{4}{n-3}\right\}\right]$$

The first term on the right gives the total number of coefficients in the nth step; the square-bracket term gives the number of conditions on the coefficients, which arise from the equations. Inside the square bracket, the first brace $\{\}$ is the number of conditions corresponding to the original equations, without taking into account identities. The second brace expresses the effect of the identities of first and second kind.

This calculation gives for Ω_n, asymptotically:

$$\Omega_n \sim \binom{4}{n}\left(\frac{18}{n} + \frac{3}{n} - \frac{24}{n} + \frac{9}{n}\right) \sim \binom{4}{n} \cdot \frac{6}{n}$$

Here we see expressed the freedom which the ψ_i have when ϕ_{ik} are given—which agrees numerically with the results obtained previously for ϕ_{ik} and ψ_i $\left(\frac{6}{n} = \frac{18}{n} - \frac{12}{n}\right)$.

In what follows, we shall consider systems of equations which are generally covariant. It is necessary here to make a remark about the number of coefficients per degree in such systems. Among the field variables let there be a tensor g_{ik}. The number of coefficients of nth degree which come from g_{ik} is then $16\binom{4}{n}$. These coefficients do not, however, all have objective meaning, since they are changed by coordinate-transformations. The

APPENDIX II

actual manifold of the truly different fields is to be judged by counting the number of independent coefficients in a *fully specialized* coordinate system.

From the transformation law it follows that, for a finite region, a coordinate choice can be made so that g_{14} , g_{24} , g_{34} vanish, and $g_{44} = \pm 1$. The remaining $g_{ik}(g_{11} \cdots g_{33})$ are then in general dependent on $x_1 \cdots x_4$. For the hypersurface $x_4 = 0$ these are functions of x_1 , x_2 , x_3 . Hence we can transform these three coordinates further so that in this hypersurface $x_4 = 0$ we have $g_{13} = g_{23} = 0$, $g_{33} = \pm 1$. This procedure can be repeated in the subspace $x_4 = x_3 = 0$, so that we get $g_{12} = 0$, $g_{22} = \pm 1$; and finally, in the subspace $x_4 = x_3 = x_2 = 0$, so that we get there $g_{11} = \pm 1$. This is the strongest possible restriction on the tensor components which can be arrived at by choice of coordinate system. As a result, we lose

$$4\binom{4}{n} + 3\binom{3}{n} + 2\binom{2}{n} + 1 = D_n$$

coefficients out of the $16\binom{4}{n}$ coefficients in the Taylor series. The general g_{ik} field has, therefore, in its nth degree, D_n less coefficients than what one would get without taking into account covariance. For large values of n we have

$$D_n \sim \binom{4}{n}\left(4 + \frac{9}{n}\right)$$

Application of the Method to the System of Equations of the Symmetric Field (Pure Gravitation)

The complete field consists of the symmetric fields g_{ik} and Γ^l_{ik} together. We consider that both fields are expanded in power series in the coordinates around the origin. The structure of the equations

$$0 = g_{ik;l} = g_{ik,l} - g_{sk}\Gamma^s_{il} - g_{is}\Gamma^s_{lk}$$

is such that in each step of differentiation the g_{ik} appear differ-

APPENDIX II

entiated one degree higher than the Γ. For that reason, we consider, in the development,

$$\Gamma = \underset{0}{\Gamma} + \underset{1}{\Gamma_s}\, x_s + \underset{2}{\Gamma_{st}}\, x_s\, x_t + \cdots$$

$\underset{0}{\Gamma}$ to be of the *first* degree, $\underset{1}{\Gamma}$ of the *second*, etc. Through this agreement we obtain the result that, in the equations which determine the coefficients of expansion, the highest coefficients of g and Γ which enter are of the same degree. The R_{ik} equations provide no difficulty, since in these the g do not enter at all.

In the nth degree of the field we have

$10\binom{4}{n}$ coefficients for the g_{ik} field

$40\binom{4}{n-1}$ coefficients for the Γ field

From these we must subtract the number D_n in order to compensate for the effect of the arbitrariness in choosing the coordinate system.

The equations $g_{ik;l} = 0$, when differentiated $n - 1$ times, give equations for the coefficients in which the highest degree of the coefficients is n. The number of such equations obtained by differentiation is

$$40\binom{4}{n-1}.$$

Equations $R_{ik} = 0$ differentiated $n - 2$ times give

$$10\binom{4}{n-2}$$

equations for coefficients of the n-degree.

However, now we must take into account the existence of four Bianchi identities, as a result of which the equations for the coefficients are not independent. The Bianchi identities give, when differentiated $n - 3$ times, identities in which the highest coefficients are of nth degree, and they appear linearly.

APPENDIX II

Hence the Bianchi identities lead to

$$4\binom{4}{n-3}$$

algebraic identities. The number of these identities must be subtracted from the total number of equations for the coefficients, which follow from the field equations; this will give us the number of independent equations for determination of the coefficients of nth degree.

Putting all this together, to subtract the number of independent equations from the number of coefficients in the nth degree, gives us the number of coefficients available for arbitrary choice. This number is

$$\Omega_n = \left[10\binom{4}{n} + 40\binom{4}{n-1} - D_n\right] - \left[40\binom{4}{n-1} + 10\binom{4}{n-2} - 4\binom{4}{n-3}\right]$$

Taking out the factor $\binom{4}{n}$, and using approximate values for large n, we get

$$\Omega_n \sim \binom{4}{n} \cdot \frac{15}{n}$$

When we compare this result with that of the wave-equation, we see that the gravitation-equations leave $\frac{5}{2}$ times as many undetermined coefficients as does the wave-equation (for $n \gg 1$). This ratio is astonishing, since the g_{ik} have ten components, while ϕ has only one.

Remark. The compatibility of the ten gravitational equations is due to the existence of the four Bianchi identities. However, it is of interest to note that the mere existence of four identities is not in itself sufficient to bring about absolute compatibility. The effect of the identities in this respect depends on the degree of differentiation in which the g_{ik} enter into this identity. If,

APPENDIX II

for instance, we had a system in which the Bianchi identities were of the *fifth*, instead of the third, degree, the calculation for Ω_n would yield asymptotically

$$\Omega_n \sim \binom{4}{n}\left(\frac{-9}{n}\right) < 0$$

which means that such a system would not be absolutely compatible.

Application of the Method to the Non-symmetric Field

Our equations are

$$\left.\begin{aligned} &\Gamma_k = 0, \qquad \overset{kl}{g}{}^{+}_{;s} = 0 \\ &R_{\underline{ik}} = 0, \qquad \underset{\vee}{R}_{ik,l} + \underset{\vee}{R}_{kl,i} + \underset{\vee}{R}_{li,k} = 0 \end{aligned}\right\} \qquad \text{(IIa)}$$

and we have six identities among them: the four Bianchi identities (of third degree of differentiation in g_{ik}); the identity $(\mathfrak{g}^{it}{}_{;t} - \mathfrak{g}^{ti}{}_{;t} - 2\mathfrak{g}^{\underline{ti}}\Gamma_t)_{,i} = 0$ (of second degree); and $\eta^{iklm}(\underset{\vee}{R}_{ik,l} + \underset{\vee}{R}_{kl,i} + \underset{\vee}{R}_{li,k})_{,m} = 0$ (of fourth degree). The number of freecoefficients is

$$\begin{aligned} \Omega_n = &\left[16\binom{4}{n} + 64\binom{4}{n-1} - D_n\right] \\ &- \left[\left\{4\binom{4}{n-1} + 64\binom{4}{n-1} + 10\binom{4}{n-2} + 4\binom{4}{n-3}\right\}\right. \\ &\left. - \left\{1\cdot\binom{4}{n-2} + 4\binom{4}{n-3} + 1\cdot\binom{4}{n-4}\right\}\right] \end{aligned}$$

or, asymptotically:

$$\Omega_n \sim \binom{4}{n}\cdot\frac{45}{n}.$$

Proof that the system-of-equations of type (II) is stronger than that of type (I), (cf. p. 142):

In system (IIa), if we could eliminate the Γ, according to

APPENDIX II

the equations $g_{ik;l} = 0$, we would find that W_{ik} (see (I) and (II)) is in that case of second degree of differentiation in the g_{ik}. In the general case the degree cannot be less; we will call it α. In case (I), the six equations $W_{ik} = 0$, lead to $\left[6\binom{4}{n-\alpha}\right]$ conditions for the coefficients in the nth degree. In case (II), the corresponding eight equations are: $\mathfrak{g}^{is}{}_{,s} = 0$, $W_{ik,l} + W_{kl,i} + W_{li,k} = 0$, with the two identities mentioned on p. 142. The number of independent conditions for system (II) is then

$$\left[4\binom{4}{n-1} + 4\binom{4}{n-\alpha-1}\right] - \left[\binom{4}{n-2} + \binom{4}{n-\alpha-2}\right]$$

When we subtract the number of conditions for system (I) from that for system (II), we get, asymptotically,

$$6\binom{4}{n} \cdot \frac{3\alpha - 4}{2n}$$

This number is positive for $\alpha \geqq 2$, which shows that system (II) is always stronger than (I).

$\left(\text{For } \alpha = 2 \text{ this difference is } \binom{4}{n} \cdot \frac{6}{n}\right)$.

§5. GENERAL REMARKS CONCERNING THE CONCEPTS AND METHODS OF THEORETICAL PHYSICS

If one does not regard as final the transition to a theory which is in principle statistical, as present-day quantum mechanics is, then the goal of a physical theory presents itself as follows: An objective (in principle complete) description of physical systems, together with a setting up of a structure of laws which connect the concepts entering into this objective description. By "objective description" is meant a description which claims potential validity and meaning without reference to any acts of observation.

Physical theories differ from mathematical structures only in the following aspect. The physical theory should provide an

APPENDIX II

essentially complete and reproducible correspondence between the conceptually described real situations and the direct sense perceptions. The question of how to set up this correspondence can only be handled intuitively, and is not expressible within the framework of the logically formulated theory.

What distinguishes one theory from another is, in the first place, the choice of its foundation stones, that is, its irreducible basic concepts out of which the structure is built.

In classical theory (mechanics) these basic concepts are the material point, the force of interaction between material points (potential energy), and the inertial system. (The latter is composed of a cartesian coordinate system together with a time coordinate.) With the increase of knowledge concerning the phenomena of the electromagnetic field, the concept of a field was added to the basic concepts on an equal footing with the material point (ponderable matter), the field being considered as a second carrier of energy.

The special theory of relativity changes this scheme only in so far as it fits into the structure of the inertial system the "fact" (really: the apparently unavoidable hypothesis based on various experiments) of the constancy of the velocity of light. The theory suggests further that we might discard the concept of the material point and deal only with the field concept. This is because the relativizing of simultaneity makes it impossible to conserve the concepts of action at a distance, and potential energy.

The modification of the theoretical foundations are altered in an intrinsically more profound way by the general theory of relativity, in that this theory does away with the inertial system altogether. In previous theories, space, represented mathematically by the inertial system, was considered to be an independent element of physical reality. This element could be thought of as absolute, inasmuch as it determined the behavior of mass points, or of the field, without itself being influenced by them. In general relativity, however, the inertial system is replaced by the displacement field, which is a component part of the total field,

APPENDIX II

this total field being the only means of description of the real world. The space aspect of real things is then completely represented by a field, which depends on four coordinate-parameters; it is a quality of this field. If we think of the field as being removed, there is no "space" which remains, since space does not have an independent existence.

So much for the *changes* which have come about in the development of these ideas. It is, however, of interest for the possibilities of future development of the theoretical foundation, to ask what is left unchanged in this development. For all these theories it is essential to operate with a space-time continuum of four dimensions (at any rate, a finite number of dimensions). (This continuum may, or may not, include singular points or lines.) The possibility of abandoning this fundamental position has often been considered, in particular by Riemann. Of course, such a renunciation of the continuum would imply a break with all the fundamental concepts of the theories considered hitherto.

All previous theories have, however one more common feature, aside from the adherence to a continuum. In all these theories it is essential that there be a group of transformations such that the same physical state is representable in different forms which go over into one another under the transformations of the group. The larger the group of transformations, the more strongly are the possible equations restricted a priori by the group.

In all these theories the (four-dimensional) *point* is considered to have objective existence, that is: existence independent of the representation. To a particular field point in *one* representation corresponds a definite point in any other equivalent representation of the same physical system. Mathematically this is expressed by the fact that only such transformations are considered equivalent as are connected by one-valued coordinate-transformations.

It is evident that the whole formalism of present day theories is inherently connected with this restriction. From it follows

APPENDIX II

the existence of a transformation law for the coordinate differential dx^i, on which is based the transformation law of a vector, and hence also of a tensor (e.g. the covariant vector can be defined by considering the invariant $\alpha_i\, dx^i$). If the concept of vector (and tensor) is lost, there is nothing which remains of the formalism of the present theories.

A priori, though, one cannot see why there could not be equivalent representations of a physical situation in which the identity of the point is not preserved (while the continuum is maintained). I have, however, no idea how to arrive at a relativistic theory in this wider sense. In any case it is essential for the understanding of the present relativity theory to know that it is based entirely on the assumption of the invariant meaning of the point.

Concluding remarks. The results of the preceding Appendix are as follows: We consider the natural generalization of the gravitational equations in empty space to be the equation-system:

$$\underset{+-}{g_{ik;s}} = 0, \qquad \Gamma_i = 0$$

$$\underset{--}{R_{ik}} = 0, \qquad \underset{\vee}{R_{ik,l}} + \underset{\vee}{R_{kl,i}} + \underset{\vee}{R_{li,k}} = 0$$

I must, however, explain why I have gone to so much trouble to arrive at this result. The contemporary physicist cannot, without such an explanation, appreciate this; for he is convinced, as a result of the successes of the probability-based quantum mechanics, that one must abandon the goal of complete descriptions of real situations in a physical theory. I do not want to discuss here why I do not share this conviction, nor why I do not consider as likely to succeed the truly interesting attempt of de Broglie and Bohm to give a complete description of the real individual situation with the formalism of the present quantum theory.

There is further the conviction that one cannot keep side by side the concepts of field and particle as elements of the physical

APPENDIX II

description. The field concept requires freedom from singularities, while the particle concept (as elementary concept) is a singularity in the field. The field concept, however, seems inevitable, since it would be impossible to formulate general relativity without it. And general relativity is the only means to avoid the "no-thing" inertial system.

For this reason I see in the present situation no possible way other than a pure field theory, which then however has before it the gigantic task of deriving the atomic character of energy. I further consider the gravitational equations of empty space to be the only rational well-founded case of a field theory, which can claim to be rigorous (also with respect to its nonlinear terms). All this leads to the attempt to arrive at a theory of a total field by generalization of the theory of gravitation for empty space.

However, setting up the generalized field equations is far from sufficient, for we know no method to find singularity-free solutions for such a system of equations—not even any method by which to judge the very existence or nonexistence of singularity-free solutions. We are thus separated by an as yet insurmountable barrier from the possibility of confronting the theory with experiment. Nevertheless, I consider it unjustified to assert, a priori, that such a theory is unable to cope with the atomistic character of energy.

APPENDIX II

RELATIVISTIC THEORY OF THE NON-SYMMETRIC FIELD

BEFORE starting with the subject proper I am first going to discuss the "strength" of systems of field equations in general. This discussion is of intrinsic interest quite apart from the particular theory presented here. For a deeper understanding of our problem, however, it is almost indispensable.

ON THE "COMPATIBILITY" AND THE "STRENGTH" OF SYSTEMS OF FIELD EQUATIONS

Given certain field variables and a system of field equations for them, the latter will not in general determine the field completely. There still remain certain free data for a solution of the field equations. The smaller the number of free data consistent with the system of field equations, the "stronger" is the system. It is clear that in the absence of any other viewpoint from which to select the equations, one will prefer a "stronger" system to a less strong one. It is our aim to find a measure for this strength of systems of equations. It will turn out that such a measure can be defined which will even enable us to compare with each other the strengths of systems whose field variables differ with respect to number and kind.

We shall present the concepts and methods involved here in examples of increasing complexity, restricting ourselves to four-dimensional fields, and in the course of these examples we shall successively introduce the relevant concepts.

APPENDIX II

First example: *The scalar wave equation**

$$\phi_{,11} + \phi_{,22} + \phi_{,33} - \phi_{,44} = 0.$$

Here the system consists of only *one* differential equation for *one* field variable. We assume ϕ to be expanded in a Taylor series in the neighborhood of a point P (which presupposes the analytic character of ϕ). The totality of its coefficients describes then the function completely. The number of nth order coefficients (that is, the nth order derivatives of ϕ at the point P) is equal to $\frac{4 \cdot 5 \ldots (n+3)}{1 \cdot 2 \ldots n}$ $\left(\text{abbreviated } \binom{4}{n}\right)$, and all these coefficients could be freely chosen if the differential equation did not imply certain relations between them. Since the equation is of second order, these relations are found by $(n-2)$ fold differentiation of the equation. We thus obtain for the nth order coefficients $\binom{4}{n-2}$ conditions. The number of nth order coefficients remaining free is therefore

$$z = \binom{4}{n} - \binom{4}{n-2}. \tag{1}$$

This number is positive for any n. Hence, if the free coefficients for all orders smaller than n have been fixed, the conditions for the coefficients of order n can always be satisfied without changing the coefficients already chosen.

Analogous reasoning can be applied to systems consisting of several equations. If the number of free nth order coefficients does not become smaller than zero, we call the system of equations *absolutely compatible*. We shall

* In the following the comma will always denote partial differentiation; thus, for example, $\phi_{,i} = \frac{\partial \phi}{\partial x^i}$, $\phi_{,11} = \frac{\partial^2 \phi}{\partial x^1 \partial x^1}$ etc.

APPENDIX II

restrict ourselves to such systems of equations. All systems known to me which are used in physics are of this kind.

Let us now rewrite equation (1). We have

$$\binom{4}{n-2} = \binom{4}{n}\frac{(n-1)n}{(n+2)(n+3)} = \binom{4}{n}\left(1 - \frac{z_1}{n} + \frac{z_2}{n^2} + \dots\right)$$

where $z_1 = +6$.

If we restrict ourselves to large values of n, we may neglect the terms $\frac{z_2}{n^2}$ etc. in the parenthesis, and we obtain for (1) *asymptotically*

$$z \sim \binom{4}{n}\frac{z_1}{n} = \binom{4}{n}\frac{6}{n}. \tag{1a}$$

We call z_1 the "coefficient of freedom," which in our case has the value 6. The larger this coefficient, the weaker is the corresponding system of equations.

Second example: *Maxwell's equations for empty space*

$$\phi^{is}{}_{,s} = 0; \qquad \phi_{ik,l} + \phi_{kl,i} + \phi_{li,k} = 0.$$

ϕ^{ik} results from the antisymmetric tensor ϕ_{ik} by raising the covariant indices with the help of

$$\eta^{ik} = \begin{pmatrix} -1 & & & \\ & -1 & & \\ & & -1 & \\ & & & +1 \end{pmatrix}.$$

These are $4 + 4$ field equations for six field variables. Among these eight equations, there exist two identities. If the left-hand sides of the field equations are denoted by G^i and H_{ikl} respectively, the identities have the form

$$G^i{}_{,i} \equiv 0; \qquad H_{ikl,m} - H_{klm,i} + H_{lmi,k} - H_{mik,l} = 0.$$

In this case we reason as follows.

APPENDIX II

The Taylor expansion of the six field components furnishes

$$6\binom{4}{n}$$

coefficients of the nth order. The conditions that these nth order coefficients must satisfy are obtained by $(n-1)$fold differentiation of the eight field equations of the first order. The number of these conditions is therefore

$$8\binom{4}{n-1}.$$

These conditions, however, are not independent of each other, since there exist among the eight equations two identities of second order. They yield upon $(n-2)$fold differentiation

$$2\binom{4}{n-2}$$

algebraic identities among the conditions obtained from the field equations. The number of free coefficients of nth order is therefore

$$z = 6\binom{4}{n} - \left[8\binom{4}{n-1} - 2\binom{4}{n-2}\right].$$

z is positive for all n. The system of equations is thus "absolutely compatible." If we extract the factor $\binom{4}{n}$ on the right-hand side and expand as above for large n, we obtain asymptotically

$$\begin{aligned} z &= \binom{4}{n}\left[6 - 8\frac{n}{n+3} + 2\frac{(n-1)n}{(n+2)(n+3)}\right] \\ &\sim \binom{4}{n}\left[6 - 8\left(1 - \frac{3}{n}\right) + 2\left(1 - \frac{6}{n}\right)\right] \\ &\sim \binom{4}{n}\left[0 + \frac{12}{n}\right]. \end{aligned}$$

APPENDIX II

Here, then, $z_1 = 12$. This shows that—and to what extent—this system of equations determines the field less strongly than in the case of the scalar wave equation ($z_1 = 6$). The circumstance that in both cases the constant term in the parenthesis vanishes expresses the fact that the system in question does not leave free any function of four variables.

Third example: *The gravitational equations for empty space.* We write them in the form

$$R_{ik} = 0; \qquad g_{ik,l} - g_{sk}\,\Gamma^s_{il} - g_{is}\,\Gamma^s_{lk} = 0.$$

The R_{ik} contain only the Γ and are of first order with respect to them. We treat here the g and Γ as independent field variables. The second equation shows that it is convenient to treat the Γ as quantities of the first order of differentiation, which means that in the Taylor expansion

$$\Gamma = \underset{0}{\Gamma} + \underset{1}{\Gamma}_s\, x^s + \underset{2}{\Gamma}_{st}\, x^s\, x^t + \ldots$$

we consider $\underset{0}{\Gamma}$ to be of the first order, $\underset{1}{\Gamma}_s$ of the second order, and so on. Accordingly, the R_{ik} must be considered as of second order. Between these equations, there exist the four Bianchi identities which, as a consequence of the convention adopted, are to be considered as of third order.

In a generally covariant system of equations a new circumstance appears which is essential for a correct enumeration of the free coefficients: fields that result from one another by mere coordinate transformations should be considered only as different representations of one and the same field. Correspondingly, only part of the

$$10\binom{4}{n}$$

nth order coefficients of the g_{ik} serves to characterize essentially different fields. Therefore, the number of expansion

APPENDIX II

coefficients that actually determine the field is reduced by a certain amount which we must now compute.

In the transformation law for the g_{ik},

$$g_{ik}^{*} = \frac{\partial x^{a}}{\partial x^{i*}} \frac{\partial x^{b}}{\partial x^{k*}} g_{ab}$$

g_{ab} and g_{ik}^{*} represent in fact the same field. If this equation is differentiated n times with respect to the x^{*}, one notices that all $(n+1)$st derivatives of the four functions x with respect to the x^{*} enter into the nth order coefficients of the g^{*}-expansion; i.e., there appear $4\binom{4}{n+1}$ numbers that have no part in the characterization of the field. In any general-relativistic theory one must therefore subtract $4\binom{4}{n+1}$ from the total number of nth order coefficients so as to take account of the general covariance of the theory. The enumeration of the free coefficients of nth order leads thus to the following result.

The ten g_{ik} (quantities of zero order of differentiation) and the forty Γ_{ik}^{l} (quantities of first order of differentiation) yield in view of the correction just derived

$$10\binom{4}{n} + 40\binom{4}{n-1} - 4\binom{4}{n+1}$$

relevant coefficients of nth order. The field equations (10 of the second and 40 of the first order) furnish for them

$$N = 10\binom{4}{n-2} + 40\binom{4}{n-1}$$

conditions. From this number, however, we must subtract

APPENDIX II

the number of the identities between these N conditions, viz.

$$4\binom{4}{n-3}$$

which result from the Bianchi identities (of the third order). Hence we find here

$$z = \left[10\binom{4}{n} + 40\binom{4}{n-1} - 4\binom{4}{n+1}\right] - \left[10\binom{4}{n-2} + 40\binom{4}{n-1}\right] + 4\binom{4}{n-3}.$$

Extracting again the factor $\binom{4}{n}$ we obtain asymptotically for large n

$$z \sim \binom{4}{n}\left[0 + \frac{12}{n}\right]. \quad \text{Thus } z_1 = 12.$$

Here, too, z is positive for all n so that the system is absolutely compatible in the sense of the definition given above. It is surprising that the gravitational equations for empty space determine their field just as strongly as do Maxwell's equations in the case of the electromagnetic field.

RELATIVISTIC FIELD THEORY

General remarks

It is the essential achievement of the general theory of relativity that it has freed physics from the necessity of introducing the "inertial system" (or inertial systems). This concept is unsatisfactory for the following reason: without any deeper foundation it singles out certain coordinate systems among all conceivable ones. It is then assumed that the laws of physics hold *only* for such inertial systems (e.g. the law of inertia and the law of the constancy

APPENDIX II

of the velocity of light). Thereby, space as such is assigned a role in the system of physics that distinguishes it from all other elements of physical description. It plays a determining role in all processes, without in its turn being influenced by them. Though such a theory is logically possible, it is on the other hand rather unsatisfactory. Newton had been fully aware of this deficiency, but he had also clearly understood that no other path was open to physics in his time. Among the later physicists it was above all Ernst Mach who focussed attention on this point.

What innovations in the post-Newtonian development of the foundations of physics have made it possible to overcome the inertial system? First of all, it was the introduction of the field concept by, and subsequent to, the theory of electromagnetism of Faraday and Maxwell, or to be more precise, the introduction of the field as an independent, not further reducible fundamental concept. As far as we are able to judge at present, the general theory of relativity can be conceived only as a field theory. It could not have developed if one had held on to the view that the real world consists of material points which move under the influence of forces acting between them. Had one tried to explain to Newton the equality of inertial and gravitational mass from the equivalence principle, he would necessarily have had to reply with the following objection: it is indeed true that relative to an accelerated coordinate system bodies experience the same accelerations as they do relative to a gravitating celestial body close to its surface. But where are, in the former case, the masses that produce the accelerations? It is clear that the theory of relativity presupposes the independence of the field concept.

The mathematical knowledge that has made it possible

APPENDIX II

to establish the general theory of relativity we owe to the geometrical investigations of Gauss and Riemann. The former has, in his theory of surfaces, investigated the metric properties of a surface imbedded in three-dimensional Euclidean space, and he has shown that these properties can be described by concepts that refer only to the surface itself and not to its relation to the space in which it is imbedded. Since, in general, there exists no preferred coordinate system on a surface, this investigation led for the first time to expressing the relevant quantities in general coordinates. Riemann has extended this two-dimensional theory of surfaces to spaces of an arbitrary number of dimensions (spaces with Riemannian metric, which is characterized by a symmetric tensor field of second rank). In this admirable investigation he found the general expression for the curvature in higher-dimensional metric spaces.

The development just sketched of the mathematical theories essential for the setting up of general relativity had the result that at first Riemannian metric was considered the fundamental concept on which the general theory of relativity and thus the avoidance of the inertial system were based. Later, however, Levi-Cività rightly pointed out that the element of the theory that makes it possible to avoid the inertial system is rather the infinitesimal displacement field Γ^l_{ik}. The metric or the symmetric tensor field g_{ik} which defines it is only indirectly connected with the avoidance of the inertial system in so far as it determines a displacement field. The following consideration will make this clear.

The transition from one inertial system to another is determined by a *linear* transformation (of a particular kind).

APPENDIX II

If at two arbitrarily distant points P_1 and P_2 there are two vectors $\underset{1}{A^i}$ and $\underset{2}{A^i}$ respectively whose corresponding components are equal to each other $(\underset{1}{A^i} = \underset{2}{A^i})$, this relation is conserved in a permissible transformation. If in the transformation formula

$$A^{i^*} = \frac{\partial x^{i^*}}{\partial x^a} A^a$$

the coefficients $\frac{\partial x^{i^*}}{\partial x^a}$ are independent of the x^a, the transformation formula for the vector components is independent of position. Equality of the components of two vectors at different points P_1 and P_2 is thus an invariant relation if we restrict ourselves to inertial systems. If, however, one abandons the concept of the inertial system and thus admits arbitrary continuous transformations of the coordinates so that the $\frac{\partial x^{i^*}}{\partial x^a}$ depend on the x^a, the equality of the components of two vectors attached to two different points in space loses its invariant meaning and thus vectors at different points can no longer be directly compared. It is due to this fact that in a general relativistic theory one can no longer form new tensors from a given tensor by simple differentiation and that in such a theory there are altogether much fewer invariant formations. This paucity is remedied by the introduction of the infinitesimal displacement field. It replaces the inertial system inasmuch as it makes it possible to compare vectors at infinitesimally close points. Starting from this concept we shall present in the sequel the relativistic field theory, carefully dispensing with anything that is not necessary to our purpose.

APPENDIX II

The infinitesimal displacement field Γ

To a contravariant vector A^i at a point P (coordinates x^t) we correlate a vector $A^i + \delta A^i$ at the infinitesimally close point $(x^t + dx^t)$ by the bilinear expression

$$\delta A^i = -\Gamma^i_{st} A^s \, dx^t \tag{2}$$

where the Γ are functions of x. On the other hand, if A is a vector field, the components of (A^i) at the point $(x^t + dx^t)$ are equal to $A^i + dA^i$ where*

$$dA^i = A^i_{,t} \, dx^t.$$

The difference of these two vectors at the neighboring point $x^t + dx^t$ is then itself a vector

$$(A^i_{,t} + A^s \Gamma^i_{st}) \, dx^t \equiv A^i_t \, dx^t$$

connecting the components of the vector field at two infinitesimally close points. The displacement field replaces the inertial system inasmuch as it effects this connection formerly furnished by the inertial system. The expression within the parenthesis, A^i_t for short, is a tensor.

The tensor character of A^i_t determines the transformation law for the Γ. We have first

$$A^i_k{}^* = \frac{\partial x^{i^*}}{\partial x^i} \frac{\partial x^k}{\partial x^{k^*}} A^i_k .$$

Using the same index in both coordinate systems is not meant to imply that it refers to corresponding components, i.e. i in x and in x^* run *independently* from 1 to 4. After

*As before "$_{,t}$" denotes ordinary differentiation $\frac{\partial}{\partial x^t}$.

APPENDIX II

some practice this notation makes the equations considerably more transparent. We now replace

$$A^i_k{}^* \text{ by } A^{i^\bullet}{}_{,k^\bullet} + A^{s^\bullet}\,\Gamma^i_{sk}{}^*$$

$$A^i_k \text{ by } A^i{}_{,k} + A^s\,\Gamma^i_{sk}$$

and again $A^{i^\bullet}$ by $\dfrac{\partial x^{i^\bullet}}{\partial x^i}A^i$, $\dfrac{\partial}{\partial x^{k^\bullet}}$ by $\dfrac{\partial x^k}{\partial x^{k^\bullet}}\cdot\dfrac{\partial}{\partial x^k}$.

This leads to an equation which, apart from the Γ^*, contains only field quantities of the original system and their derivatives with respect to the x of the original system. Solving this equation for the Γ^* one obtains the desired transformation formula

$$\Gamma^i_{kl}{}^* = \frac{\partial x^{i^\bullet}}{\partial x^i}\frac{\partial x^k}{\partial x^{k^\bullet}}\frac{\partial x^l}{\partial x^{l^\bullet}}\Gamma^i_{kl} - \frac{\partial^2 x^{i^\bullet}}{\partial x^s \partial x^l}\frac{\partial x^s}{\partial x^{k^\bullet}}\frac{\partial x^l}{\partial x^{l^\bullet}} \tag{3}$$

whose second term (on the right-hand side) can be somewhat simplified:

$$-\frac{\partial^2 x^{i^\bullet}}{\partial x^s \partial x^l}\frac{\partial x^s}{\partial x^{k^\bullet}}\frac{\partial x^l}{\partial x^{l^\bullet}}$$

$$= -\frac{\partial}{\partial x^{l^\bullet}}\left(\frac{\partial x^{i^\bullet}}{\partial x^s}\right)\frac{\partial x^s}{\partial x^{k^\bullet}} = -\frac{\partial}{\partial x^{l^\bullet}}\left(\frac{\partial x^{i^\bullet}}{\partial x^{k^\bullet}}\right)\cdot\cdot + \frac{\partial x^{i^\bullet}}{\partial x^s}\frac{\partial^2 x^s}{\partial x^{k^\bullet}\partial x^{l^\bullet}}$$

$$= \frac{\partial x^{i^\bullet}}{\partial x^s}\frac{\partial^2 x^s}{\partial x^{k^\bullet}\partial x^{l^\bullet}} \tag{3a}$$

We call such a quantity a *pseudo tensor*. Under linear transformations it transforms as a tensor, whereas for non-linear transformations a term is added which does not contain the expression to be transformed, but only depends on the transformation coefficients.

APPENDIX II

Remarks on the displacement field.

1. The quantity $\tilde{\Gamma}^i_{kl}$ ($\equiv\Gamma^i_{lk}$) which is obtained by a transposition of the lower indices also transforms according to (3) and is therefore likewise a displacement field.

2. By symmetrizing or anti-symmetrizing equation (3) with respect to the lower indices k^*, l^* one obtains the two equations

$$\Gamma^i_{\underline{kl}}{}^* \left(= \tfrac{1}{2}(\Gamma^i_{kl}{}^* + \Gamma^i_{lk}{}^*)\right) = \frac{\partial x^{i^*}}{\partial x^i}\frac{\partial x^k}{\partial x^{k^*}}\frac{\partial x^l}{\partial x^{l^*}}\Gamma^i_{\underline{kl}} - \frac{\partial^2 x^{i^*}}{\partial x^s \partial x^t}\frac{\partial x^s}{\partial x^{k^*}}\frac{\partial x^t}{\partial x^{l^*}}$$

$$\Gamma^i_{\underset{\smile}{kl}}{}^* \left(= \tfrac{1}{2}(\Gamma^i_{kl}{}^* - \Gamma^i_{lk}{}^*)\right) = \frac{\partial x^{i^*}}{\partial x^i}\frac{\partial x^k}{\partial x^{k^*}}\frac{\partial x^l}{\partial x^{l^*}}\Gamma^i_{\underset{\smile}{kl}}\,.$$

Hence the two (symmetric and anti-symmetric) constituents of Γ^i_{kl} transform independently of each other, i.e. without mixing. Thus they appear from the point of view of the transformation law as independent quantities. The second of the equations shows that $\Gamma^i_{\underset{\smile}{kl}}$ transforms as a tensor. From the point of view of the transformation group it seems therefore at first unnatural to combine these two constituents additively into one single quantity.

3. On the other hand, the lower indices of Γ play quite different roles in the defining equation (2) so that there is no compelling reason to restrict the Γ by the condition of symmetry with regard to the lower indices. If one does so nevertheless, one is led to the theory of the pure gravitational field. If, however, one does not subject the Γ to a restrictive symmetry condition, one arrives at that generalization of the law of gravitation that appears to me as the natural one.

The curvature tensor

Although the Γ-field does not itself have tensor character, it implies the existence of a tensor. The latter is most

APPENDIX II

easily obtained by displacing a vector A^i according to (2) along the circumference of an infinitesimal two-dimensional surface element and computing its change in one circuit. This change has vector character.

Let $\overset{0}{x}{}^l$ be the coordinates of a fixed point and x^l those of another point on the circumference. Then $\xi^l = x^l - \overset{0}{x}{}^l$ is small for all points of the circumference and can be used as a basis for the definition of orders of magnitude.

The integral $\oint \delta A^i$ to be computed is then in more explicit notation

$$-\oint \underline{\Gamma^i_{st}}\, \underline{A^s}\, dx^t \quad \text{or} \quad -\oint \underline{\Gamma^i_{st}}\, \underline{A^s}\, d\xi^t.$$

Underlining of the quantities in the integrand indicates that they are to be taken for successive points of the circumference (and not for the initial point, $\xi^l = 0$).

We first compute in the lowest approximation the value of $\underline{A^i}$ for an arbitrary point ξ^l of the circumference. This lowest approximation is obtained by replacing in the integral, extended now over an open path, $\underline{\Gamma^i_{st}}$ and $\underline{A^s}$ by the values Γ^i_{st} and A^s for the initial point of integration ($\xi^l = 0$). The integration gives then

$$\underline{A^i} = A^i - \Gamma^i_{st} A^s \int d\xi^t = A^i - \Gamma^i_{st} A^s \xi^t.$$

What is neglected here, are terms of second or higher order in ξ. With the same approximation one obtains immediately

$$\underline{\Gamma^i_{st}} = \Gamma^i_{st} + \Gamma^i_{st,r}\, \xi^r.$$

APPENDIX II

Inserting these expressions in the integral above one obtains first, with an appropriate choice of the summation indices,

$$-\oint (\Gamma^i_{st} + \Gamma^i_{st,q}\,\xi^q)\,(A^s - \Gamma^s_{pq}\,A^p\xi^q)\,d\xi^t$$

where all quantities, with the exception of ξ, have to be taken for the initial point of integration. We then find

$$-\Gamma^i_{st}\,A^s \oint d\xi^t - \Gamma^i_{st,q}\,A^s \oint \xi^q\,d\xi^t + \Gamma^i_{st}\,\Gamma^s_{pq}\,A^p \oint \xi^q\,d\xi^t$$

where the integrals are extended over the closed circumference. (The first term vanishes because its integral vanishes.) The term proportional to $(\xi)^2$ is omitted since it is of higher order. The two other terms may be combined into

$$[-\Gamma^i_{pt,q} + \Gamma^i_{st}\,\Gamma^s_{pq}]A^p \oint \xi^q\,d\xi^t.$$

This is the change ΔA^i of the vector A^i after displacement along the circumference. We have

$$\oint \xi^q\,d\xi^t = \oint d(\xi^q\xi^t) - \oint \xi^t\,d\xi^q = -\oint \xi^t\,d\xi^q.$$

This integral is thus antisymmetric in t and q, and in addition it has tensor character. We denote it by f^{tq}. If f^{tq} were an *arbitrary* tensor, then the vector character of ΔA^i would imply the tensor character of the bracketed expression in the last but one formula. As it is, we can only infer the tensor character of the bracketed expression if antisymmetrized with respect to t and q. This is the *curvature tensor*

$$R^i_{klm} \equiv \Gamma^i_{kl,m} - \Gamma^i_{km,l} - \Gamma^i_{sl}\,\Gamma^s_{km} + \Gamma^i_{sm}\,\Gamma^s_{kl}. \qquad (4)$$

The position of all indices is hereby fixed. Contracting

APPENDIX II

with respect to i and m one obtains the *contracted curvature tensor*

$$R_{ik} \equiv \Gamma^s_{ik,s} - \Gamma^s_{is,k} - \Gamma^s_{it}\,\Gamma^t_{sk} + \Gamma^s_{ik}\,\Gamma^t_{st} \tag{4a}$$

The λ-transformation

The curvature has a property which will be important in the sequel. For a displacement field Γ we may define a new Γ^* according to the formula

$$\Gamma^l_{ik}{}^* = \Gamma^l_{ik} + \delta^l_i \lambda_{,k} \tag{5}$$

where λ is an arbitrary function of the coordinates, and δ^l_i is the Kronecker tensor ("λ-transformation"). If one forms $R^i_{klm}(\Gamma^*)$ by replacing Γ^* by the right-hand side of (5), λ cancels so that

$$\left.\begin{aligned} R^i_{klm}(\Gamma^*) &= R^i_{klm}(\Gamma) \\ R_{ik}\ (\Gamma^*) &= R_{ik}\ (\Gamma) \end{aligned}\right\} \tag{6}$$

and

The curvature is invariant under λ-transformations ("λ-invariance"). Consequently, a theory which contains Γ only within the curvature tensor cannot determine the Γ-field completely but only up to a function λ, which remains arbitrary. In such a theory, Γ and Γ^* are to be regarded as representations of the same field, in the same way as if Γ^* were obtained from Γ merely by a coordinate transformation.

It is noteworthy that the λ-transformation, contrary to a coordinate transformation, produces a non-symmetric Γ^* from a Γ that is symmetric in i and k. The symmetry condition for Γ loses in such a theory its objective significance.

The main significance of λ-invariance lies in the fact that it has an influence on the "strength" of the system of the field equations, as we shall see later.

APPENDIX II

The requirement of "transposition invariance"

The introduction of non-symmetric fields meets with the following difficulty. If Γ^l_{ik} is a displacement field, so is $\tilde{\Gamma}^l_{ik}$ $(= \Gamma^l_{ki})$. If g_{ik} is a tensor, so is $\tilde{g}_{ik}$ $(= g_{ki})$. This leads to a large number of covariant formations among which it is not possible to make a selection on the principle of relativity alone. We shall demonstrate this difficulty by an example and we shall show how it can be overcome in a natural manner.

In the theory of the symmetric field the tensor

$$(W_{ikl} \equiv)\ g_{ik,l} - g_{sk}\,\Gamma^s_{il} - g_{is}\,\Gamma^s_{lk}$$

plays an important part. If it is put equal to zero, one obtains an equation which permits to express the Γ by the g, i.e. to eliminate the Γ. Starting from the facts that (1) $A^i_{\ t} \equiv A^i_{\ ,t} + A^s\Gamma^i_{st}$ is a tensor, as proved earlier, and that (2) an arbitrary contravariant tensor can be expressed in the form $\sum_{t\ (t)\ (t)} A^i B^k$, it can be proved without difficulty that the above expression has tensor character also if the fields g and Γ are no longer symmetric.

But in the latter case, the tensor character is not lost if, e.g., in the last term Γ^s_{lk} is transposed, i.e. replaced by Γ^s_{lk} (this follows from the fact that g_{is} $(\Gamma^s_{kl} - \Gamma^s_{lk})$ is a tensor). There are other formations, though not quite so simple, that conserve the tensor character and can be regarded as extensions of the above expression to the case of the non-symmetric field. Consequently, if one wants to extend to non-symmetric fields the relation between the g and the Γ obtained by setting the above expression equal to zero, this seems to involve an arbitrary choice.

But the above formation has a property that distinguishes it from the other possible formations. If one replaces in

APPENDIX II

it simultaneously g_{ik} by $\tilde{g}_{ik}$ and Γ^l_{ik} by $\tilde{\Gamma}^l_{ik}$ and then interchanges the indices i and k it is transformed into itself: it is "transposition symmetric" with respect to the indices i and k. The equation obtained by putting this expression equal to zero is "transposition invariant." If g and Γ are symmetric, this condition is, of course, also satisfied; it is a generalization of the condition that the field quantities be symmetric.

We postulate for the field equations of the non-symmetric field that they be *transposition invariant.* I think that this postulate, physically speaking, corresponds to the requirement that positive and negative electricity enter symmetrically into the laws of physics.

A glance at (4a) shows that the tensor R_{ik} is not completely transposition symmetric, since it transforms by transposition into

$$(R_{ik}^* =)\ \Gamma^s_{ik,s} - \Gamma^s_{sk,i} - \Gamma^s_{it}\,\Gamma^t_{sk} + \Gamma^s_{ik}\,\Gamma^t_{ts}\,. \qquad (4b)$$

This circumstance is the basis of the difficulties that one encounters in the endeavour to establish transposition invariant field equations.

The pseudo tensor U^l_{ik}

It turns out that a transposition symmetric tensor can be formed from R_{ik} by the introduction of a somewhat different pseudo tensor U^l_{ik} instead of Γ^l_{ik}. In (4a) the two terms that are linear in Γ can be formally combined to a single one. One replaces $\Gamma^s_{ik,s} - \Gamma^s_{is,k}$ by $(\Gamma^s_{ik} - \Gamma^t_{it}\,\delta^s_k)_{,s}$ and defines a new pseudo tensor U^l_{ik} by the equation

$$U^l_{ik} \equiv \Gamma^l_{ik} - \Gamma^t_{it}\,\delta^l_k\,. \qquad (7)$$

Since

$$U^t_{it} = -3\,\Gamma^t_{it}\,,$$

APPENDIX II

as follows from (7) by contraction with respect to k and l, we obtain the following expression for Γ in terms of U:

$$\Gamma^l_{ik} = U^l_{ik} - \tfrac{1}{3}\, U^t_{it}\, \delta^l_k\,. \tag{7a}$$

Inserting these in (4a) one finds

$$S_{ik} \equiv U^s_{ik,s} - U^s_{it}\, U^t_{sk} + \tfrac{1}{3}\, U^s_{is}\, U^t_{tk} \tag{8}$$

for the contracted curvature tensor in terms of U. This expression, however, is transposition symmetric. It is this fact that makes the pseudo tensor U so valuable for the theory of non-symmetric fields.

λ-transformation for U. If in (5) the Γ are replaced by the U, one obtains by a simple calculation

$$U^l_{ik}{}^* = U^l_{ik} + (\delta^l_i\, \lambda_{,k} - \delta^l_k\, \lambda_{,i})\,. \tag{9}$$

This equation defines the λ-transformation for the U. (8) is invariant with respect to this transformation ($S_{ik}(U^*) = S_{ik}(U)$).

The transformation law for U. If in (3) and (3a) the Γ are replaced by the U with the help of (7a), one obtains

$$U^l_{ik}{}^* = \frac{\partial x^{l^*}}{\partial x^l}\,\frac{\partial x^i}{\partial x^{i^*}}\,\frac{\partial x^k}{\partial x^{k^*}}\,U^l_{ik} + \frac{\partial x^{l^*}}{\partial x^s}\,\frac{\partial^2 x^s}{\partial x^{i^*}\partial x^{k^*}} - \delta^{t^*}_{k^*}\,\frac{\partial x^{t^*}}{\partial x^s}\,\frac{\partial^2 x^s}{\partial x^{i^*}\partial x^{t^*}}\,. \tag{10}$$

Note that again the indices referring to both systems assume all the values from 1 to 4 *independently of each other*, even though the same letter is being used. Regarding this formula it is noteworthy that on account of the last term it is not transposition symmetric with respect to the indices i and k. This peculiar circumstance can be clarified by demonstrating that this transformation may be regarded as a composition of a transposition symmetric coordinate

APPENDIX II

transformation and a λ-transformation. In order to see that we write first the last term in the form

$$-\tfrac{1}{2}\left[\delta^{l^*}_{k^*}\frac{\partial x^{t^*}}{\partial x^s}\frac{\partial^2 x^s}{\partial x^{i^*}\partial x^{t^*}}+\delta^{l^*}_{i^*}\frac{\partial x^{t^*}}{\partial x^s}\frac{\partial^2 x^s}{\partial x^{k^*}\partial x^{t^*}}\right]$$
$$+\tfrac{1}{2}\left[\delta^{l^*}_{i^*}\frac{\partial x^{t^*}}{\partial x^s}\frac{\partial^2 x^s}{\partial x^{k^*}\partial x^{t^*}}-\delta^{l^*}_{k^*}\frac{\partial x^{t^*}}{\partial x^s}\frac{\partial^2 x^s}{\partial x^{i^*}\partial x^{t^*}}\right].$$

The first of these two terms is transposition symmetric. Let us combine it with the first two terms of the right-hand side of (10) to an expression $K^l_{ik}{}^*$. Let us now consider what we get if the transformation

$$U^l_{ik}{}^* = K^l_{ik}{}^*$$

is followed by the λ-transformation

$$U^l_{ik}{}^{**} = U^l_{ik}{}^* + \delta^{l^*}_{i^*}\lambda_{,k^*} - \delta^{l^*}_{k^*}\lambda_{,i^*}.$$

The composition yields

$$U^l_{ik}{}^{**} = K^l_{ik}{}^* + (\delta^{l^*}_{i^*}\lambda_{,k^*} - \delta^{l^*}_{k^*}\lambda_{,i^*}).$$

This implies that (10) may be regarded as such a composition provided the second term of (10a) can be brought into the form $\delta^{l^*}_{i^*}\lambda_{,k^*} - \delta^{l^*}_{k^*}\lambda_{,i^*}$. For this it is sufficient to show that a λ exists such that

$$\tfrac{1}{2}\frac{\partial x^{t^*}}{\partial x^s}\frac{\partial^2 x^s}{\partial x^{k^*}\partial x^{t^*}} = \lambda_{,k^*} \qquad (11)$$

$$\left(\text{and } \tfrac{1}{2}\frac{\partial x^{t^*}}{\partial x^s}\frac{\partial^2 x^s}{\partial x^{i^*}\partial x^{t^*}} = \lambda_{,i^*}\right).$$

In order to transform the left-hand side of the so far hypothetical equation we must first express $\dfrac{\partial x^{t^*}}{\partial x^s}$ by the coefficients of the inverse transformation, $\dfrac{\partial x^a}{\partial x^{b^*}}$. On the one hand,

$$\frac{\partial x^p}{\partial x^{t^*}}\frac{\partial x^{t^*}}{\partial x^s} = \delta^p_s. \qquad (a)$$

APPENDIX II

On the other,

$$\frac{\partial x^p}{\partial x^{t^\bullet}} V^s_{t^\bullet} = \frac{\partial x^p}{\partial x^{t^\bullet}} \frac{\partial D}{\partial \left(\frac{\partial x^s}{\partial x^t}\right)} = D\delta^p_s.$$

Here, $V^s_{t^\bullet}$ denotes the co-factor of $\frac{\partial x^s}{\partial x^{t^\bullet}}$, and may in turn be expressed as the derivative of the determinant $D = \left|\frac{\partial x^a}{\partial x^{b^\bullet}}\right|$ with respect to $\frac{\partial x^s}{\partial x^{t^\bullet}}$. Therefore, we have also

$$\frac{\partial x^p}{\partial x^{t^\bullet}} \cdot \frac{\partial \log D}{\partial \left(\frac{\partial x^s}{\partial x^{t^\bullet}}\right)} = \delta^p_s. \qquad \text{(b)}$$

It follows from (a) and (b) that

$$\frac{\partial x^{t^\bullet}}{\partial x^s} = \frac{\partial \log D}{\partial \left(\frac{\partial x^s}{\partial x^{t^\bullet}}\right)}.$$

Because of this relation the left-hand side of (11) can be written as

$$\tfrac{1}{2} \frac{\partial \log D}{\partial \left(\frac{\partial x^s}{\partial x^{t^\bullet}}\right)} \left(\frac{\partial x^s}{\partial x^{t^\bullet}}\right)_{,k^\bullet} = \tfrac{1}{2} \frac{\partial \log D}{\partial x^{k^\bullet}}.$$

This implies that (11) is indeed satisfied by

$$\lambda = \tfrac{1}{2} \log D.$$

This proves that the transformation (10) can be regarded as a composition of the transposition symmetric transformation

$$U^l_{ik}{}^* = \frac{\partial x^{l^\bullet}}{\partial x^l} \frac{\partial x^i}{\partial x^{i^\bullet}} \frac{\partial x^k}{\partial x^{k^\bullet}} U^l_{ik} + \frac{\partial x^{l^\bullet}}{\partial x^s} \frac{\partial^2 x^s}{\partial x^{i^\bullet} \partial x^{k^\bullet}} - \tfrac{1}{2} \left[\delta^{l^\bullet}_{k^\bullet} \frac{\partial x^{t^\bullet}}{\partial x^s} \frac{\partial^2 x^s}{\partial x^{i^\bullet} \partial x^{t^\bullet}} + \delta^{l^\bullet}_{i^\bullet} \frac{\partial x^{t^\bullet}}{\partial x^s} \frac{\partial^2 x^s}{\partial x^{k^\bullet} \partial x^{t^\bullet}} \right] \qquad \text{(10b)}$$

APPENDIX II

and a λ-transformation. (10b) may thus be taken in place of (10) as transformation formula for the U. Any transformation of the U-field that only changes the *form* of the representation can be expressed as a composition of a coordinate transformation according to (10b) and a λ-transformation.

Variational principle and field equations

The derivation of the field equations from a variational principle has the advantage that the compatibility of the resulting system of equations is assured and that the identities connected with the general covariance, the "Bianchi identities," as well as the conservation laws result in a systematic manner.

The integral to be varied requires as integrand $\mathfrak{H}$ a scalar density. We shall construct such a density from R_{ik} or S_{ik}. The simplest procedure is to introduce a covariant tensor density $\mathfrak{g}^{ik}$ of weight 1 in addition to Γ or U respectively, setting

$$\mathfrak{H} = \mathfrak{g}^{ik} R_{ik} \ (= \mathfrak{g}^{ik} S_{ik}). \tag{12}$$

The transformation law for the $\mathfrak{g}^{ik}$ must be

$$\mathfrak{g}^{i^*k^*} = \frac{\partial x^{i^*}}{\partial x^i} \frac{\partial x^{k^*}}{\partial x^k} \mathfrak{g}^{ik} \left| \frac{\partial x^l}{\partial x^{l^*}} \right| \tag{13}$$

where again the indices referring to different coordinate systems, in spite of the use of the same letters, are to be treated as independent of each other. We obtain indeed

$$\int \mathfrak{H}^* \, d\tau^* = \int \frac{\partial x^{i^*}}{\partial x^i} \frac{\partial x^{k^*}}{\partial x^k} \mathfrak{g}^{ik} \left| \frac{\partial x^l}{\partial x^{l^*}} \right| \cdot \frac{\partial x^s}{\partial x^{i^*}} \frac{\partial x^t}{\partial x^{k^*}} S_{st} \left| \frac{\partial x^{r^*}}{\partial x^r} \right| d\tau$$

$$= \int \mathfrak{H} \, d\tau,$$

APPENDIX II

i.e. the integral is transformation invariant. Furthermore, the integral is invariant with respect to a λ-transformation (5) or (9) because R_{ik} as expressed by the Γ or U respectively, and hence also $\mathfrak{H}$, is invariant with respect to a λ-transformation. From this it follows that also the field equations to be derived by variation of $\int \mathfrak{H} d\tau$ are covariant with respect to coordinate and to λ-transformations.

But we also postulate that the field equations are to be transposition invariant with respect to the two fields $\mathfrak{g}$, Γ or the fields $\mathfrak{g}$, U. This is assured if $\mathfrak{H}$ is transposition invariant. We have seen that R_{ik} is transposition symmetric if expressed in the U, but not if expressed in the Γ. Hence $\mathfrak{H}$ is only transposition invariant if we introduce in addition to the $\mathfrak{g}^{ik}$ the U (but not the Γ) as field variables. In that case, we are sure from the beginning that the field equations derived from $\int \mathfrak{H}\, d\tau$ by variation of the field variables are transposition invariant.

By variation of $\mathfrak{H}$ (equations (12) and (8)) with respect to the $\mathfrak{g}$ and U we find

$$\left.\begin{aligned} \delta\mathfrak{H} &= S_{ik}\,\delta\mathfrak{g}^{ik} - \mathfrak{N}^{ik}{}_{l}\,\delta U^{l}_{ik} + (\mathfrak{g}^{ik}\,\delta U^{s}_{ik})_{,s} \\ \text{where}\quad S_{ik} &= U^{s}_{ik,s} - U^{s}_{il}\,U^{t}_{sk} + \tfrac{1}{3}\,U^{s}_{is}\,U^{t}_{tk}, \\ \mathfrak{N}^{ik}{}_{l} &= \mathfrak{g}^{ik}{}_{,l} + \mathfrak{g}^{sk}\,(U^{i}_{sl} - \tfrac{1}{3}\,U^{t}_{st}\,\delta^{i}_{l}) \\ &\quad + \mathfrak{g}^{is}\,(U^{k}_{ls} - \tfrac{1}{3}\,U^{t}_{ts}\,\delta^{k}_{l}). \end{aligned}\right\} \qquad (14)$$

The field equations

Our variational principle is

$$\delta\left(\int \mathfrak{H}\, d\tau\right) = 0 \qquad (15)$$

APPENDIX II

The $\mathfrak{g}^{ik}$ and U^l_{ik} are to be varied independently, their variations vanishing at the boundary of the domain of integration. This variation gives first of all

$$\int \delta\mathfrak{H}\, d\tau = 0.$$

If the expression given in (14) is inserted here, the last term of the expression for $\delta\mathfrak{H}$ does not give any contribution since δU^l_{ik} vanishes at the boundary. Hence we obtain the field equations

$$S_{ik} = 0 \tag{16a}$$

$$\mathfrak{N}^{ik}{}_l = 0. \tag{16b}$$

They are—as is already evident from the choice of the variational principle—invariant with respect to coordinate and to λ-transformations and also transposition invariant.

Identities

These field equations are not independent of each other. Between them exist $4 + 1$ identities. That is, there exist $4 + 1$ equations between their left-hand sides that hold regardless of whether or not the $\mathfrak{g}$-U field satisfies the field equations.

These identities can be derived by a well-known method from the fact that $\int \mathfrak{H}\, d\tau$ is invariant with respect to coordinate and to λ-transformations.

For it follows from the invariance of $\int \mathfrak{H}\, d\tau$ that its variation vanishes *identically* if one inserts in $\delta\mathfrak{H}$ the variations $\delta\mathfrak{g}$ and δU which arise from an infinitesimal coordinate transformation or an infinitesimal λ-transformation respectively.

APPENDIX II

An infinitesimal coordinate transformation is described by

$$x^{i^*} = x^i + \xi^i \tag{17}$$

where ξ^i is an arbitrary infinitesimal vector. We must now express the δg^{ik} and δU^l_{ik} by the ξ^i using the equations (13) and (10b). Because of (17) one must replace

$$\frac{\partial x^{a^*}}{\partial x^b} \text{ by } \delta^a_b + \xi^a{}_{,b}, \quad \frac{\partial x^a}{\partial x^{b^*}} \text{ by } \delta^a_b - \xi^a{}_{,b}$$

and omit all terms that are of higher than first order in ξ. Thus one obtains

$$\delta g^{ik} (= g^{ik^*} - g^{ik}) = g^{sk}\,\xi^i{}_{,s} + g^{is}\,\xi^k{}_{,s} - g^{ik}\,\xi^s{}_{,s} + [-g^{ik}{}_{,s}\,\xi^s] \tag{13a}$$

$$\delta U^l_{ik} (= U^{l*}_{ik} - U^l_{ik}) = U^s_{ik}\,\xi^l{}_{,s} - U^l_{sk}\,\xi^s{}_{,i} - U^l_{is}\,\xi^s{}_{,k} + \xi^l{}_{,ik} + [-U^l{}_{ik,s}\,\xi^s]. \tag{10c}$$

Note here the following. The transformation formulas furnish the new values of the field variables *for the same point of the continuum*. The calculation indicated above first gives expressions for δg^{ik} and δU^l_{ik} without the terms in brackets. In the calculus of variation, on the other hand, δg^{ik} and δU^l_{ik} denote the variations *for fixed values of the coordinates*. In order to obtain these the terms in brackets have to be added.

If one inserts in (14) these "transformation variations" δg and δU, the variation of the integral $\int \mathfrak{H}\, d\tau$ vanishes identically. If furthermore the ξ^i are so chosen that they vanish together with their first derivatives at the boundary of the domain of integration, the last term in (14) gives no contribution. The integral

$$\int (S_{ik}\,\delta g^{ik} - \mathfrak{N}^{ik}{}_l\,\delta U^l_{ik})\, d\tau$$

APPENDIX II

vanishes therefore identically if the δg^{ik} and δU^l_{ik} are replaced by the expressions (13a) and (10c). Since this integral depends linearly and homogeneously on the ξ^i and their derivatives, it can be brought into the form

$$\int \mathfrak{W}_i \, \xi^i \, d\tau$$

by repeated integration by parts, where $\mathfrak{W}_i$ is a known expression (of first order in S_{ik} and of second order in $\mathfrak{N}^{ik}_l$). From this follow the identities

$$\mathfrak{W}_i \equiv 0. \tag{18}$$

These are four identities for the left-hand sides S_{ik} and $\mathfrak{N}^{ik}_l$ of the field equations, which correspond to the Bianchi identities. According to the terminology introduced before these identities are of third order.

There exists a fifth identity corresponding to the invariance of the integral $\int \mathfrak{H} \, d\tau$ with respect to infinitesimal λ-transformations. Here we have to insert in (14)

$$\delta g^{ik} = 0 \qquad \delta U^l_{ik} = \delta^l_i \lambda_{,k} - \delta^l_k \lambda_{,i}$$

where λ is infinitesimal and vanishes at the boundary of the domain of integration. One obtains first

$$\int \mathfrak{N}^{ik}_l \, (\delta^l_i \lambda_{,k} - \delta^l_k \lambda_{,i}) \, d\tau = 0$$

or, after integration by parts,

$$2 \int \mathfrak{N}^{is}_{s,i} \, \lambda \, d\tau = 0$$

(where, generally, $\mathfrak{N}^{ik}_l = \frac{1}{2} (\mathfrak{N}^{ik}_l - \mathfrak{N}^{ki}_l)$).

APPENDIX II

This furnishes the desired identity

$$\mathfrak{N}^{is}_{\underset{\sim}{s},i} \equiv 0. \tag{19}$$

In our terminology this is an identity of second order. For $\mathfrak{N}^{is}_{\underset{\sim}{s}}$ we obtain from (14) by straightforward computation

$$\mathfrak{N}^{is}_{\underset{\sim}{s}} \equiv \mathfrak{g}^{is}_{\underset{\sim}{,s}}. \tag{19a}$$

If the field equation (16b) is satisfied, we have thus

$$\mathfrak{g}^{is}_{\underset{\sim}{,s}} = 0. \tag{16c}$$

Remark on the physical interpretation. A comparison with Maxwell's theory of the electromagnetic field suggests the interpretation that (16c) expresses the vanishing of the magnetic current density. If this is accepted, it is evident which expression should denote the electric current density. One can assign a tensor g^{ik} to the tensor density $\mathfrak{g}^{ik}$ by setting

$$\mathfrak{g}^{ik} = g^{ik}\sqrt{-|g_{st}|} \tag{20}$$

where the covariant tensor g_{ik} is correlated to the contravariant one by the equations

$$g_{is}\,g^{ks} = \delta^k_i. \tag{21}$$

From these two equations we obtain

$$g^{ik} = \mathfrak{g}^{ik}\,(-|\mathfrak{g}^{st}|)^{-\frac{1}{2}}$$

and then g_{ik} from equations (21). We may then assume that

$$(a_{ikl}) = g_{\underset{\sim}{ik},l} + g_{\underset{\sim}{kl},i} + g_{\underset{\sim}{li},k} \tag{22}$$

or

$$\mathfrak{a}^m = \tfrac{1}{6}\,\eta^{iklm}\,a_{ikl} \tag{22a}$$

expresses the current density, where η^{iklm} is Levi-Cività's tensor density (with components ± 1) antisymmetric in all indices. The divergence of this quantity vanishes identically.

APPENDIX II

The strength of the system of equations (16*a*), (16*b*)

In applying here the method of enumeration described above one must take into account the fact that all the U^* obtained from a given U by λ-transformations of the form (9) actually represent the same U-field. This has the consequence that the nth order coefficients of the U^l_{ik}-expansion incorporate $\binom{4}{n}$ nth order derivatives of λ whose choice is of no consequence for the distinction of actually differing U-fields. Thus the number of expansion coefficients relevant for the enumeration of the U-fields is decreased by $\binom{4}{n}$. By the enumeration method we obtain for the number of free nth order coefficients

$$z = \left[16\binom{4}{n} + 64\binom{4}{n-1} - 4\binom{4}{n+1} - \binom{4}{n}\right] - \left[16\binom{4}{n-2} + 64\binom{4}{n-1}\right] + \left[4\binom{4}{n-3} + \binom{4}{n-2}\right]. \tag{23}$$

The first bracket represents the total number of relevant nth order coefficients which characterize the g-U-field, the second the reduction of this number due to the existence of the field equations, and the third bracket gives the correction to this reduction on account of the identities (18) and (19). Computing the asymptotic value for large n we find

$$z \sim \binom{4}{n}\frac{z_1}{n} \tag{23a}$$

where

$$z_1 = 42.$$

APPENDIX II

The field equations of the non-symmetric field are thus considerably weaker than those of the pure gravitational field ($z_1 = 12$).

The influence of λ-invariance on the strength of the system of equations. One may be tempted to bring about transposition invariance of the theory by starting from the transposition invariant expression

$$\mathfrak{H} = \tfrac{1}{2}(\mathfrak{g}^{ik} R_{ik} + \tilde{\mathfrak{g}}^{ik} \tilde{R}_{ik})$$

(instead of introducing the U as field variables). Of course, the resulting theory will be different from the one expounded above. It can be shown that for this $\mathfrak{H}$ no λ-invariance exists. Here, too, we obtain field equations of the type (16a), (16b), which are transposition invariant (with respect to $\mathfrak{g}$ and Γ). Between them, there exist, however, only the four "Bianchi identities." If one applies the method of enumeration to this system, then, in the formula corresponding to (23), the fourth term in the first bracket and the second term in the third bracket are missing. One obtains

$$z_1 = 48.$$

The system of equations is thus weaker than the one chosen by us and is therefore to be rejected.

Comparison with the previous system of field equations. This is given by

$$\Gamma^s_{\underset{\smile}{is}} = 0 \qquad R_{\underline{ik}} = 0$$

$$g_{ik,l} - g_{sk}\,\Gamma^s_{il} - g_{is}\,\Gamma^s_{lk} = 0 \qquad R_{\underset{\smile}{ik},l} + R_{\underset{\smile}{kl},i} + R_{\underset{\smile}{li},k} = 0$$

where R_{ik} is defined by (4a) as a function of the Γ (and where $R_{\underline{ik}} = \frac{1}{2}(R_{ik} + R_{ki})$, $R_{\underset{\smile}{ik}} = \frac{1}{2}(R_{ik} - R_{ki})$).

This system is entirely equivalent to the new system

APPENDIX II

(16a), (16b) since it has been derived from the same integral by variation. It is transposition invariant with respect to the g_{ik} and Γ^l_{ik}. The difference, however, lies in the following. The integral to be varied is itself not transposition invariant, nor is the system of equations that is at first obtained by its variation; it is, however, invariant with respect to the λ-transformations (5). In order to obtain transposition invariance here, one has to use an artifice. One formally introduces four new field variables, λ_i, which after variation are so chosen that the equations $\Gamma^s_{\underset{\smile}{is}} = 0$ are satisfied.* Thus the equations obtained by variation with respect to the Γ are brought into the indicated transposition invariant form. But the R_{ik}-equations still contain the auxiliary variables λ_i. One can, however, eliminate them, which leads to a decomposition of these equations in the manner stated above. The equations obtained are then also transposition invariant (with respect to the $\mathfrak{g}$ and Γ).

Postulating the equations $\Gamma^s_{\underset{\smile}{is}} = 0$ involves a normalization of the Γ-field, which removes the λ-invariance of the system of equations. As a result, not all equivalent representations of a Γ-field appear as solutions of this system. What takes place here, is comparable to the procedure of adjoining to the field equations of pure gravitation arbitrary additional equations which restrict the choice of coordinates. In our case, moreover, the system of equations becomes unnecessarily complicated. These difficulties are avoided in the new representation by starting from a variational principle that is transposition invariant with respect to the $\mathfrak{g}$ and U, and by using throughout the $\mathfrak{g}$ and U as field variables.

* By setting $\Gamma^l_{ik}{}^* = \Gamma^l_{ik} + \delta^l_i \lambda_k$.

APPENDIX II

The divergence law and the conservation law of momentum and energy

If the field equations are satisfied and if, moreover, the variation is a transformation variation, then, in (14), not only S_{ik} and $\mathfrak{N}^{ik}_{l}$ vanish, but also $\delta\mathfrak{H}$, so that the field equations imply the equations

$$(\mathfrak{g}^{ik}\,\delta U^{s}_{ik})_{,s} = 0$$

where δU^{s}_{ik} is given by (10c). This divergence law holds for any choice of the vector ξ^{i}. The simplest special choice, i.e. ξ^{i} independent of the x, leads to the four equations

$$\mathfrak{T}^{s}_{l,s} \equiv (\mathfrak{g}^{ik}\,U^{s}_{ik,l})_{,s} = 0.$$

These can be interpreted and applied as the equations of conservation of momentum and energy. It should be noted that such conservation equations are never uniquely determined by the system of field equations. It is interesting that according to the equations

$$\mathfrak{T}^{s}_{l} \equiv \mathfrak{g}^{ik}\,U^{s}_{ik,l}$$

the density of the energy current $(\mathfrak{T}^{1}_{4}, \mathfrak{T}^{2}_{4}, \mathfrak{T}^{3}_{4})$ as well as the energy density $\mathfrak{T}^{4}_{4}$ vanish for a field that is independent of x^{4}. From this one can conclude that according to this theory a stationary field free from singularities can never represent a mass different from zero.

The derivation as well as the form of the conservation laws become much more complicated if the former formulation of the field equations is used.

GENERAL REMARKS

A. In my opinion the theory presented here is the logically simplest relativistic field theory which is at all

APPENDIX II

possible. But this does not mean that nature might not obey a more complex field theory.

More complex field theories have frequently been proposed. They may be classified according to the following characteristic features:

(a) Increase of the number of dimensions of the continuum. In this case one must explain why the continuum is *apparently* restricted to four dimensions.

(b) Introduction of fields of a different kind (e.g. a vector field) in addition to the displacement field and its correlated tensor field g_{ik} (or $\mathfrak{g}^{ik}$).

(c) Introduction of field equations of higher order (of differentiation).

In my view, such more complicated systems and their combinations should be considered only if there exist physical-empirical reasons to do so.

B. A field theory is not yet completely determined by the system of field equations. Should one admit the appearance of singularities? Should one postulate boundary conditions? As to the first question, it is my opinion that singularities must be excluded. It does not seem reasonable to me to introduce into a continuum theory points (or lines etc.) for which the field equations do not hold. Moreover, the introduction of singularities is equivalent to postulating boundary conditions (which are arbitrary from the point of view of the field equations) on "surfaces" which closely surround the singularities. Without such a postulate the theory is much too vague. In my opinion the answer to the second question is that the postulation of boundary conditions is indispensable. I shall demonstrate this by an elementary example. One can compare the postulation

APPENDIX II

of a potential of the form $\phi = \sum \frac{m}{r}$ with the statement that outside the mass points (in three dimensions) the equation $\Delta\phi = 0$ is satisfied. But if one does not add the boundary condition that ϕ vanish (or remain finite) at infinity, then there exist solutions that are entire functions of the x (e.g. $x_1^2 - \frac{1}{2}(x_2^2 + x_3^2)$) and become infinite at infinity. Such fields can only be excluded by postulating a boundary condition in case the space is an "open" one.

C. Is it conceivable that a field theory permits one to understand the atomistic and quantum structure of reality? Almost everybody will answer this question with "no." But I believe that at the present time nobody knows anything reliable about it. This is so because we cannot judge in what manner and how strongly the exclusion of singularities reduces the manifold of solutions. We do not possess any method at all to derive systematically solutions that are free of singularities. Approximation methods are of no avail since one never knows whether or not there exists to a particular approximate solution an exact solution *free of singularities*. For this reason we cannot at present compare the content of a nonlinear field theory with experience. Only a significant progress in the mathematical methods can help here. At the present time the opinion prevails that a field theory must first, by "quantization," be transformed into a statistical theory of field probabilities according to more or less established rules. I see in this method only an attempt to describe relationships of an essentially nonlinear character by linear methods.

D. One can give good reasons why reality cannot at all be represented by a continuous field. From the quantum phenomena it appears to follow with certainty that a finite

APPENDIX II

system of finite energy can be completely described by a finite set of numbers (quantum numbers). This does not seem to be in accordance with a continuum theory, and must lead to an attempt to find a purely algebraic theory for the description of reality. But nobody knows how to obtain the basis of such a theory.

PART IV

THE POPULAR LECTURES

1
INTRODUCTORY REMARKS

The first two lectures, given in Princeton on 9 and 10 May 1921, under the title "Generalities on the Theory of Relativity (Without Using Mathematical Symbols)," were attended by a large, nonprofessional audience that almost filled the lecture hall. The first lecture was dedicated to the special theory of relativity, and the second presented the transition to the general theory and discussed the experimental tests of its validity.

The lectures were delivered in German and a German stenographer took shorthand notes. The typescript of these notes was not edited and is reproduced, in its rudimentary form, in CPAE, vol. 7, as Appendix C. The stenographer left spaces in the manuscript enclosed by square brackets whenever words or formulas were incomprehensible. Blank lines and half-lines indicate that whole phrases or sentences seem to be missing. The lectures were concluded by an oral summary in English by Professor Edwin Adams. His summaries were published the following day in the *New York Evening Post.* They accompany the German texts in the Collected Papers of Albert Einstein.

A press report in the *Daily Princetonian* on Einstein's first lecture reads:

> Dr. Einstein delivered his lecture in his native tongue, President Hibben presiding. The hall was nearly full, and it was difficult for those in the rear to hear the speaker who spoke in rather low voice. After the talk, Prof. E. P. Adams of the University Faculty, gave a resume in English of the scientist's remarks.[1]

The *Evening Bulletin* (Philadelphia) described the speaker in vivid picturesque terms:

> The scientist picked up a piece of chalk before he began to talk. Part of the theory he explained on the blackboard. The rest he explained out in space, with the chalk drawing imaginary lines. The long hair that ended in tight curls and the chalk balanced between his fingers like the baton gave him the appearance of an orchestra leader. . . . He wore a green knitted tie with a thin red stripe in it, perfectly creased trousers and a formal black coat.[2]

Journalists reporting on the second lecture grasped the importance that Einstein attached to the as yet unconfirmed prediction of his theory: the shift of the spectral lines toward the red by the gravitational field of the sun (the gravitational redshift). They preceded their reports with sensational headline statements, like the one in the *Daily Princetonian*:

EINSTEIN SAYS THEORY IS LIABLE TO COLLAPSE

Permanency Depends upon Validity of Conclusion Regarding Vibration from Sun—Bent Spectral Lines Will Prove or Disprove—Unless They Lean Slightly Toward Red of the Spectrum, Relativity is a Fallacy[3]

We present here, for the first time, an English translation of these lectures. The translator, Dieter Brill, to whom we are grateful for his effort and for the result of his effort, had to cope with dilemmas and face difficult choices in performing this task. We asked him to describe the translation process (see box).

NOTE FROM THE TRANSLATOR

The pages I translated are described as "A transcription of the typescript of the two popular lectures, made from notes taken by a stenographer at the lectures." At each of these transformations of Einstein's spoken word there was opportunity to introduce errors or omissions. I tried to preserve the meaning of Einstein's words while correcting errors from previous stages and avoiding introductions of new errors. The most striking omissions in the transcription are passages denoted by blank lines, hyphens (----) or brackets [ ]. The longer omissions may be due to diagrams or equations that the stenographer could not reproduce, but I expect that most of them were occasions where the stenographer could not follow due to speed of speech or unfamiliarity with the subject. These omissions were retained in the translation, except in cases where it seemed quite clear what word or phrase was omitted, which was then placed in the originally blank bracket. I have also used the bracket notation for insertions where the transcript had no gap. In a few cases, my insertion is not unique and only indicates the grammatical structure of an incomplete sentence, which would otherwise be ambiguous due to lack of inflection in English. I have also tried to correct errors that are likely due to misunderstandings by the stenographer. These changes are documented in footnotes, where phrases in quotation marks are the German of the transcription, and English translations are in parentheses. Finally, I have broken up some exceedingly long sentences, and added paragraph breaks to make the lectures easier to read.

DIETER BRILL, MARCH 2016.

Very minor edits have been made to the text. One must bear in mind that during the lectures Einstein used a blackboard to write mathematical expressions and sometimes to draw figures to illustrate his thoughts. The transcript of the stenographer's notes does not reflect this and therefore in some places the text is difficult to follow. This is the case, for example, when Einstein derives the transformation between different coordinate systems (the Lorentz transformation) or when he discusses the perihelion precession of Mercury. Nevertheless, the text as it stands reflects the vivaciousness of Einstein's presentation and the enthusiasm of his delivery.

In his first two lectures, Einstein explains very thoroughly the principles and conceptual foundations of, first, special and then general relativity with a minimum of mathematical apparatus. He also includes qualitative discussions of the main experiential evidence supporting his theories, such as the Michelson-Morley experiment (which Einstein calls the Michelson experiment) as support for the validity of the relativity principle, and also of the propagation of light. The guiding theme is the principle of relativity, which he begins to explain in basic kinematic terms. The lectures in some sense recapitulate Einstein's own intellectual pathway in that they review, step-by-step, the contradictions he encountered and the resolutions he found for them. In the conclusion to his first lecture, he emphasizes that the most important task of a physical theory is to reduce the number of independent assumptions rooted in experience. He also stressed that the unification of the conservation laws for mass and energy is the most important result of special relativity. In the conclusion to his second lecture, he emphasizes that according to general relativity, one can no longer first construe a region of space and time and then place objects therein, but rather that spatial and temporal aspects of reality are inevitably tied up with all other physical laws.

NOTES

1. József Illy, ed. *Albert Meets America: How Journalists Treated Genius during Einstein's 1921 Travels* (Baltimore: Johns Hopkins Univ. Press, 2006), pp. 169–70.
2. Quoted in ibid., p. 172.
3. Quoted in ibid., p. 179.

2
EINSTEIN'S LECTURES

FIRST LECTURE

[9 May 1921]

Ladies and Gentlemen!

I will arrange my lectures about relativity in this way: today I present first some general ideas, either without any, or with a minimum of formal mathematical support, and in the later lectures I will expand somewhat more explicitly on the question of what the basic idea of the theory of relativity is. The theory of relativity is so named because this whole theory is concerned with the question of the extent to which any motion is merely relative motion. Namely, the point of it is this. When we generally speak of the motion of a body, then by the concept of motion we always mean relative motion. For example, when we speak of a car moving in the street, the motion refers to the piece of ground or surface called the road, and this piece of Earth's surface plays the role of a body on which this motion will develop. Thus, the very idea of motion is relative motion, and according to the concept of motion one could equally well say the street moves relative to the car, as we can say the car moves relative to the street. These are quite self-evident circumstances, and in a way have nothing to do with the theory of relativity. Yet the fact is that we can conceive of motion only as relative motion. [C]oncerning the pure geometrical acceleration it is indifferent from which body [one talks about it].

That is self-evident and needs no detailed elaboration. But the question arises whether motion is absolutely relative also from the point of view of physics. Well, what does that mean? This question has, for example, the following meaning: One can ask whether there is a preferred state of motion in the world that plays such a fundamental role with respect to [ ] that one can say, this state of motion is tantamount to absolute rest. In the absence of one preferred state of motion in the world, one can pose the question whether there is a group of preferred states of motion. This conception is of the kind that was established in classical mechanics by Galileo and Newton. But the theory of relativity addresses the question whether or not [all] the physical processes by their regular structure are so constituted that there are preferred states of motion with respect to these natural processes. Well, the theory of relativity consists of two separate and distinct parts, the first part is called the "special" theory of relativity, and the second the "general" theory of relativity. Like classical mechanics, the special theory

of relativity claims that there is not only one preferred state of motion, but an infinite variety of preferred states of motion. The general theory of relativity goes far beyond this standpoint and claims that the fundamental laws are such that preferred states of motion do not exist at all in the world.

Of course, these general words say very little yet, and therefore I want to show you how the questions of concern were developed step by step. I can easily show you this by first taking a short look at the classical mechanics of Galileo and Newton. The principle at the foundation of classical mechanics is the law of inertia established by Galileo. This law of inertia says that motion is uniform and rectilinear for a body that is sufficiently distant from other bodies so that these other bodies do not act on it. So the motion of [such] a special point is a rectilinear, uniform motion. This law of G[alileo] already provides a hint that the laws of mechanics know no absolutely preferred motion. Namely, there is the following: We have already seen previously that as soon as we speak of motion, we must avail ourselves of a coordinate system, that is, of a reference body to which the motion relates. I will assume that a particular coordinate system is used to describe the motion, and we will assume that the theory of relativity and Galileo's law of inertia are valid. Let this for example be the trajectory of a special point that is sufficiently distant from other points. Now one can easily see that if this motion is related to a new reference body that has some other state of motion, e.g., a new reference body [K'] that is in uniform rotation relative to K [ ], then the law of inertia is no longer valid relative to K'. For, simple geometrical reasoning shows that the trajectory of a point left to its own devices, which is a straight-line, uniform motion in reference to K, is no straight-line, uniform motion relative to the rotating K'. So you see that G[alileo]'s law of inertia must properly be [reformulated]. One should say: there are coordinate systems in the world, or at least one coordinate system in the world, of such states of motion that[1] a mass point sufficiently removed from other mass points appears in a straight-line, uniform motion relative to that coordinate system. The motion is straight-line and uniform if it is[2] referred to a K system. Now we want to show that according to mechanics there is not only one coordinate system K, or one particular state of motion for the coordinate system K, but infinitely many. We imagine in addition to the reference body a second reference body that moves rectilinearly and uniformly relative to K. So besides K, the system relative to which the law of inertia holds, there shall exist yet a second uniform and non-rotating system, which we will again call K'. Then a simple geometrical consideration again shows that the same body that [moves uniformly in K] also moves[3] uniformly relative to K'. That means that the law of inertia is equally valid relative to K' as it is relative to K, and generally we can say that if the foundation of G[alileo-]N[ewtonian] mechanics is of such a nature that, given a coordinate system K relative to which these basic laws hold, then these same basic laws hold relative to every other coordinate system that moves rectilinearly and uniformly compared to the original [coordinate] system. Thus you see that the states of motion of all such coordinate systems K/K' are equivalent to each other, at least in regard to the basic laws of G[alileo-]N[ewtonian] mechanics.

We now want to introduce a name for the presumption that the laws hold not only for K, but also for every uniformly moving coordinate system. This assertion that the laws are valid independent of the choice of the coordinate system, if only such coordinate systems are considered that move rectilinearly and uniformly against each other, this we call the special relativity principle. So this principle states that natural laws that are

valid relative to K also [are valid in K'] if K' moves uniformly relative to K, and of course the question arises, is this special relativity principle a general law of nature or is it valid only for G[alileo-]N[ewtonian] mechanics? Now, the development of electricity theory and optical has taught us to believe that this relativity principle is no generally valid law, but that also among the so-constituted coordinate systems, which we also call inertial systems, that is, [it appears] that for the electro-mechanical and [optical processes] there is not K and K', but that for electr[icity] there is a single preferred coordinate system, so that the special relativity principle enjoys a kind of accidental validity only within the mechanical phenomena known to us, but that in the universe there would be one preferred state of motion. We ask[4], how did physicists come to believe such a thing? Experience did not force us directly into such a belief, but the theories, which developed on the basis of experience, seem[ed] to show that there must be a preferred coordinate system in the universe. Namely, it was principally the phenomena of the interference of light and the refraction of light that mandated the concept of light as a kind of wave process, a [ ] process. The ideas of that time, formed in the first, and partially in the second half of the nineteenth century, were that all natural processes must be understood in the last instance as processes of motion, so it was natural that the (orientation) propagation of light was also construed as a state of motion, and in order to comprehend such a motion mechanically one needed a medium that moves. And now one asks, what is the nature of this medium, which propagates light. And the polarization properties compelled the interpretation of light oscillations as transverse[5], and that already implied that one would have to regard the light medium that one had in mind as a kind of solid body, for only these have the property of propagating transverse waves. Now for rapid oscillations a body could be, so to speak, solid, and physicists therefore asked whether this [was the] state partially or not. And so it turned out in particular from the electro-magnetic research of L[orentz?] that almost all phenomena can be interpreted effortlessly by the assumption that the light-aether does not participate in the motion of bodies, that is, that the aether is to be conceived as something whose parts in the whole universe are at rest relative to each other. And naturally connected with this was of course the necessary image that a preferred state of motion also exists in the universe, similar to that of the light-medium. Let us identify that state of motion with the coordinate system K. And if this understanding is correct, namely that there is a preferred state of motion in the universe, then from what was said about this preferred K, and from to Maxwell's equations of electricity, and from all accumulated experience, one could say, the laws of propagation in this system K are so designed that every light ray spreads out in every direction with the speed 300,000 km, whereby this propagation speed shall be independent of whatever is the state of motion assumed by the light-infused body. This can be [considered] a safe consequence from the totality of theoretical [physics] and this principle plays a very special role for the development of the theory of relativity.[6] For short, we will call this a principle of constant-speed.

1. The relativity principle
2. The light principle

These are two assumptions discovered indirectly from experience, and now reasoning appears to teach us the following. These two assumptions, the relativity principle and the principle of constancy of the speed of light, appear to contradict each other, in the

following sense in particular. If it is true that the principle of constancy of the speed of light prevails, then it seems that the relativity principle must be invalid for electro-magnetic and optical phenomena. This seeming contradiction between these two principles can be easily displayed by means of a simple drawing. If a light ray has speed C relative to the system K then, it is claimed, it cannot have the same speed C relative to the uniformly moving system K'. Let us imagine a light ray [is] sent off [from] an origin toward the right through empty space, and we want to see where this light ray ends up when one second is over. Then this light ray makes relative to the [ ] that means the light has the speed C. Now let us assume that

* * *

Now we can say, the distance of the light front from the origin via coordinate system K' is C−V, according to the figure. The difference ----

So the conclusion is this, that a light ray, which propagates in this direction and is observed from K', would have speed [C−V]; had we considered a light ray, which goes in the opposite direction, then we would have [C+V] and our conclusion is, relative to K all [light rays] go [at C], relative to K' a different law of light propagation applies, for the light propagates [at C±V]. That means the principle of constancy of the speed of light is not valid in reference to K', the relativity principle does not apply to the light. Well, for some time no one doubted this [violation] of the relativity principle at all, it was the conviction of physicists that there [is] a [preferred system] in reference to which the laws of nature must find their natural expression. The laws of nature in their natural and simple formulation need a coordinate system of [ ]. Due to the development of optics, the conviction took hold that there is a physically preferred state of motion in the universe, namely the one in reference to which the law of constancy of the speed of light is valid. But there are several very accurate experiments that argue against this view. Namely, if it is true that there is a preferred state of motion in the universe, and that the optical phenomena, viewed from a coordinate system in a different state of motion, depend on direction, if that is true, then it must be possible to capture this through experience. After all, here we are sitting on the earth making physical experiments and because the earth runs around the sun with a speed of 30 K[m/sec] it is clear that we cannot be sitting the whole year at rest in the K [system]. So to speak we sit on the earth in a system K' and it is clear that relative to the earth the law of constancy of light propagation cannot be valid. This is the basis of the famous experiment of Michelson, which was in principle as follows. Imagine a rigid rod and let a ray of light move along this rigid rod, and let us say this rigid rod is mounted lengthwise in the direction of the earth's motion. One can easily calculate how long the light ray needs to go back and forth along this rod, if it is in motion relative to [K]. The calculation shows that this time is not exactly equal to the time that light would need to go back and forth along this rod if this rod were at rest. So the fact is that if this rod takes part in the motion with the earth relative to the light aether, the light needs a longer time to go back and forth than if the rod were at rest. We can now ask how large is the round trip time if the same rod lies perpendicular, and there it turns out that this time has a somewhat different value.

* * *

By this it would be shown that in regard to light propagation, space [would be anisotropic as seen] from K', that is, [seen] from the earth it would have different properties in different directions, due to the motion of the earth relative to the light-aether. The experiment of Michelson [has tested this] by an optical method of enormous precision. This time difference does not exist. So light takes as long to propagate back and forth perpendicular to the direction of motion [as along that direction]. So it appears that the consequence derived from the principle of constancy of the light's <speed> of propagation is not right, rather it appears that also for light the special relativity principle can claim validity, as far as experience goes. And now a fatal contradiction follows. The optical phenomena are understandable only on the basis of the Maxwell-[Lorentz] equations. On the other hand, the Michelson experiment shows that, contrary to this aether theory, the special relativity principle appears to be valid also when applied to light. Now the question is, how can this contradiction be put aside. And so it turned out that this contradiction owes to the inclusion of arbitrary and unproven premises in the previously described interpretation,[7] and it turned out that a careful analysis of the processes show that the relativity principle and the principle of constancy of the speed of light are not incompatible, as seems to be the case.

* * *

If my claim is correct, there must be a mistake or an arbitrary presumption imbedded in this [reasoning], and this mistake resp. arbitrary presumption is what I must now exhibit, and therein consists the essential idea of the special theory of relativity, and therefore I must insist on calling attention to this point. It turns out that in all physical considerations we have made one arbitrary presumption about the physical notion of time, and about the notion of simultaneity in particular. There was never any doubt that it makes objective sense when two events that take place at two different locations are said to be simultaneous. Let us consider two stars somewhere that [both] suddenly flare up while moving in an arbitrary way. The belief was that the event of flare-up of these two stars was simultaneous, and the aim of the present claim is that this is an erroneous presumption. In reality, without a specific definition one knows nothing at all when assuming that two such events are simultaneous. Our accepted opinion of what constitutes simultaneous phenomena follows from our everyday experience that light gives us information about events that take place at a certain distance from us with such speed that we believe, with the help of light, to directly experience simultaneity. If, for example, electric lamps light up at both ends of this hall, and an observer situated in this part of the hall takes note of the simultaneity of the events, then another observer, located in the other part of the hall reports the same, because light represents for us a blitz communication medium, and this makes the concept "simultaneous" physically self-evident to us. But if the events take place at very distant locations from each other, then the concept of simultaneity is no longer evident. We can imagine that two physicists somewhere in the world are in dispute about two such events, whether these were simultaneous or not; they cannot settle their dispute because they do not know what they are saying as long as no definition exists on the basis of which it can be established whether the events are simultaneous or not; and so we have to give a definition. We want to make the definition as follows. We will say: we have the fundamental belief that, upon appropriate formulation of the natural laws, the law of constancy of light propagation is correct, and since we are also

convinced of the correctness of the special relativity principle, so we will be permitted to base the definition of simultaneity on just these two principles. We proceed thus: we use some inertial system or other, I will consider it marked by this rigid rod, which shall be oriented so that the two events that interest us take place immediately next to it. Then I say, I determine the locations where the events take place relative to this inertial system, which I may designate by K, and set up the inertial system between these two points. I can achieve this by repeated application of a measuring rod. ----

Now if the two light signals from the two events arrive simultaneously at an observer who is located in the middle, then we say they are simultaneous. Here we are defining "simultaneous" by means of the theorem about light propagation. Of course it would be possible that this definition leads to contradictions, for the concept of simultaneity involves the doctrine that if an event A is simultaneous with an event B, and simultaneous with an event C, then the events B & C are also simultaneous, and we know ---- But we will assume it, for we have indeed defined simultaneity, namely by means of the theorem about the constancy of light propagation. For this definition we need an inertial system K. We could also have used a [ ] system "I" and, insofar as we maintain the special relativity principle, we must say, all these K [systems] are equivalent by their nature. So we must say that this definition of simultaneity does not define simultaneity per se, but only simultaneity in relation to a chosen inertial system. With this simultaneity becomes a contingent concept, which belongs to an "I" system to which we refer. It is also easy to see by a simple argument that if [K] and K' are coincident with "I", which are moving uniformly relative to each other, then events that are simultaneous in reference to K are not simultaneous in reference to the "I" system K'. So we see that if we want to maintain this, in accordance with experience, then we must in exchange abandon the doctrine that simultaneity is an absolute concept. But until now it was a tacit assumption in physics that simultaneity has absolute meaning. You will easily see from our argument above, that the propagation of light does contain the premise that time has a meaning independent of the coordinate system. For we were saying we consider the spreading of the light in one second, and the light gets to this point here

* * *

and we said that this is how far light spreads out in one second. But now we must say that we have no right whatsoever to claim that light, as in this case ---- So with this we have found an arbitrary step in the previous argument, and the contradiction between these two premises becomes invalid. Yet another arbitrary premise is contained [in the argument]. We postulated that the length is C−V. Also from [the point of view of] coordinate K' ----

But this is in no way self-evident. The situation is namely this: think of a rod here, which moves longitudinally with speed V relative to the blackboard and now we ask, how long is this rod. About that we can say, viewed from the rod itself we get the length, let us say a unit rod, which we denote by [ ]. This is the length of the rod as viewed from the rod itself. But if we want to distance ourselves from the blackboard [then] we must take a instantaneous photograph of the rod [and] we must draw the points of the rod on the blackboard where they [ ] a ----
and if we choose these two, then we can indicate on the blackboard the difference between these two points. When we have done this, we see that these two procedures, measuring from the rod and from the blackboard, are totally different in their definition, and the

question is now whether these two manipulations lead to the same result. So we cannot know [ ]. We presupposed that ---- is not a length in relation to the system K but also [ ] and this is the second arbitrary premise of the argument. And now that we drop these two arbitrary premises, it can very well be that the special relativity principle and the principle of constancy of the speed of light are re[conciled] and the mathematical analysis shows that this re[conciliation] indeed exists. As long as the two arbitrary premises, absolute meaning of simultaneity and absolute meaning of length without dependence on the motion, were given, it was easy to solve the following problem in a purely analytical way. Let any general processes [be] referred to an inertial system K, that is, all events that we examine physically [are referred to] by knowing their spatial coordinates ----

If this is known then it is [just] a question of calculation what are now the coordinates and time values of all these events we study with respect to a second coordinate system. So we ask ----

By way of a very simple coordinate deduction one finds its transformation equations, which are based on the premise of ----
and the independence of the duration of the motion. But when these premises are dropped, then one realizes that one is not immediately able to solve the problem. We are no longer equipped offhand to find out by pure calculation what the space and what the time coordinates are of an event if the relevant XYZT are known in respect to the coordinate system K. This transformation equation cannot be immediately derived if those arbitrary premises are not made. Don't we have at hand the means to solve this problem given the two principles, derived from experience, namely the principle of the constancy of light propagation and the relativity principle? Before doing this I want briefly to say something about the meaning of these XYZT ----

The coordinate axes[8] are to be considered as rigid bodies and the ---- are to be taken as constructed with the help of rigid bodies. To determine the length of these coordinates one needs a unit measuring rod, which we have to consider as given once and for all. We imagine many copies of such a measuring rod, which have the same length compared to each other, and with them these distances can be measured. We also imagine identical clocks erected at the different points, unit clocks, and these shall be so adjusted that equal positions of the hands are simultaneous; they shall be regulated with the help of the principle of the constancy of the speed of light. When these identical clocks are adjusted in this way, then they indicate the time at their location. The essence in this matter is that the space and time readings we use have a direct physical meaning because they are obtained [ ] with the help of rigid bodies and of clocks, and we have to keep this in mind. So, when we begin with space and time coordinates, we are dealing with results of measurements, which are obtained with identical measuring rods and clocks, which are in the state of rest in respect to the two coordinates X resp. X'. If we now of the several events XY are valid, then we learn something about the behavior of clocks that are at rest relative to K', that is, they are in motion relative to K, how these behave when viewed from K. These designations are not just purely formal in nature, rather they make a statement about the physical behavior of measuring rods and clocks. How can we know transformation laws? These transformation laws must be able to specify how the principle about constancy of the speed of light is valid with respect to K as well as to K'. When you formulate mathematically the conditions that the relation between space and time of a

system K and another inertial system shall be such that the principle about constancy of the speed of light is valid with respect to both inertial systems, then it turns out that the transformation laws of space and time between K and K', that is, the relations that exist between [their coordinates] ---- are of a well-defined type. In particular, in the case that the coordinate systems are so oriented relative to each other that they have a common X-axis and that they ---- then these equations take on forms like ones used already by Lorentz in his investigations. This general type of transformation[9] equation represents for me the central point of the formal theory of the special theory of relativity. I will try later to explicate for you to what extent that is the case. These equations express that the spatial-temporal relations between the two coordinate systems, and that the law for the speed of light applies to both, and between these two postulates there is no contradiction whatsoever, otherwise there would not be such equations. We say these equations [ ] that the metric relations between the two coordinate systems. ---- if we assess this/these from a coordinate system that does not participate in the motion. It turns out that a rod of length L [in] a state of longitudinal motion . . . which when measured at rest [has] a certain length, and if it is referred to this space, that is, to a coordinate system that does not participate in the motion, that then compared to the previous[10] result of this measurement [yields] a [smaller length] than at rest. So it has a smaller length than in the state of rest.

* * *

and the rod is shortened longitudinally. [Whereas] a rod moving transverse to its extent does not experience contraction. A body that has the form of a sphere when observed at rest and that we now set in motion with speed V relative to our coordinate system, stays a sphere for a co-moving observer. Viewed from us it becomes flattened. If it were possible to move the body itself with the speed of light, the body would be completely turned into a flat disk. It is impossible to introduce a speed of a coordinate system which reaches the speed C, for at V ---- and the transformation equations lose their sense, since for physical motion the speed of light plays the role of a limiting speed, which cannot be exceeded. If we introduce two systems besides K, a system K' and a second K" and we assume that K' moves relative to K at less than the speed of light, and if further we choose the third system so that it, relative to [K' moves at C]. We ask, how large is the speed of [K"], then we can calculate this easily with the aid of the transformation equations. It turns out [ ], so that by addition of velocities it is never possible to reach the speed of light, but that the speed of light plays the role of a physical limiting speed. This shows up when considering the extent of moving bodies.

The second consequence is the following. A clock, which indicates each second correctly when observed at rest relative to the coordinate system, runs more slowly when it is put in a state of motion, in fact, the slow-down is again given by the root [ ]. These two consequences appear to be strange, but logically that is how it is. Should one of these consequences not be right, then either the law about the propagation of light or the special relativity principle must be false, and we shall cite strong reasons from experience for both of these laws. In contrast, about the shape of measuring rods or the [running] speed of clocks that move at such high speed that this speed becomes non-negligible compared to the speed of light, we know nothing. Therefore we must retain these equations. Having

these transformation equations we can make use of them more extensively than merely to find out how moving measuring rods and clocks behave. Further reasoning will have to go like this. If it is true that all inertial systems are equivalent for the formulation of the laws [ ] of nature, then the following must be the case. We think of all natural laws as being established by experience and
we think of the laws determined also in reference to K'. Then we can say the following. If the special relativity law is correct, then the laws in reference to K' must be exactly the same as in reference to K, or with this difference that here of course, when we substantiate in respect to K', there occur the symbols [ ], which are defined in K. But now we do have the mathematical connection that exists between the coordinates of K and K', for this mathematical connection is given by the Lorentz transformation. In a purely mathematical way, we can determine the natural laws in respect to K' if these are known in respect to K. We only need to carry out a transformation, that is, to introduce [ ]. If then the special relativity principle is to be valid in the form demanded by these equations, then upon transformation we must [come] to exactly the same [ ] as were valid in K. Thus transformation of [the laws] with the help of the Lorentz transformation [must] lead [to the laws] in reference to K'. This is a condition that the natural laws must fulfill if it is to be true that the special relativity principle is valid for them. We can therefore say: the special relativity principle contains a general mathematical condition, which the natural laws have to satisfy.

There is in this regard[11], epistemologically, a [parallellism][12] between the theory of relativity and ther[modynamics]. Namely, ther[modynamics] proceeds as follows. How must the laws of nature be constituted so that one cannot construct with their help [perpetual motion machines], and one draws the mathematical consequences from the condition that the laws of nature are such that one cannot with their help construct a [perpetual motion machine]. Here we ask, how must the natural laws be so that with their help

[one] can prefer neither, so that in respect to the formulation of the natural laws they are fully equivalent. The natural laws must be so constructed that upon conversion, with their help, they are transformed into laws of concordant structure.

Now the question still remains: what has this theory of relativity now delivered by way of definite results? These can be found by applying those transformations to the particular ones. Mathematics has found better methods, and it is not necessary to engage in them now. How does a physical law have to look when one applies this transformation to that law? How must equations be set up so that they [ ] with the help of these equations? The two most important equations we had at that time for representing natural phenomena were the equations of the electro-

So it appears that they are such that in equivalent [ ] they [ ]. This is no wonder, for the law of constancy of the speed of light is effectively a [ ] of Maxwell's. As long as this [theory] was not so well developed, one could not claim that the law about the speed of light [ ] the mechanical [one] is shown not to be easily fulfilled, according to the special theory of relativity. This is understandable, namely, according to Newton's law of motion, there is no principle difficulty in accelerating a body to arbitrarily high speed, even to speeds that are greater than the speed of light C. Whereas we have already seen that according to special relativity re[lations] the speed of light C plays the role of a physical limit for speed. The modifications, which the equations undergo, are of this type, that instead of

[ ] the proper derivation of the motional [ ]. The whole difference consists of this, that for the impact of a body, the momentum it has, which enters in ordinary mechanics through a speed-limiting [square] root, [ ]. In [ ] this has been expressed this way: the mass of a body increases with speed. The mass as an inertial property takes the role of a [rest mass] in this connection. It is a quantity that does not change during a motion. By their nature these [square] roots belong rather with the velocity [components]. Thus, it comes about that no[13] speeds like the speed of light can occur. These equations, which [ ] motions of very fast moving particles [ ] from the electrical [ ] will thereby represent this with high accuracy. The main achievement of the special theory of relativity rests on this, that it substantially decreases the independent premises of electro-dynamics, namely by making it possible to reduce all those laws, referring to temporal change in a process, to the [transformation] laws, as[14] one substantially reduces the independent [ ] of the theory. It is the highest task of the [ ] [that one] reduce the [axioms] of the theory to a minimum with as little as possible [ ] on each other, apart from this, that they are based on two assumptions[15] that were forced on us by experience.

The physical [ ] consequence provided by the special theory of relativity refers to the nature of mass. Before the special theory of relativity, one had the view that in the universe there are two quite specially important laws of conservation for matter, namely, firstly the law of conservation of mass, and secondly the law of conservation of energy. These two laws seem[ed] to be independently valid. But now one can show, using the methods of the special theory of relativity, that the law of conservation of mass acquires a limitation, so that it together

Like all general laws of nature, so also the law of conservation of energy must comply with the relativity principle. If one now formulates this, that the law of conservation of matter be valid in respect to K as well as in respect to K', one arrives at the following consequence. Assuming one has an arbitrary system of bodies with total mass M, that is, with an inertia given by the number M, and one imparts to this system an amount of energy E then the inertial mass of this body is increased, to be specific [by E/c^2], so by addition of this energy E the mass becomes [$M + E/c^2$]. As long as the internal processes taking place in a system work in such a way that they do not change the energy of the system, the mass stays the same, but as soon as the energy changes, the mass changes as well and one can say mass is nothing but a measure for the energy. The measures of those main parts [ ] which cannot be essentially changed. This insight that the [ ] is the main result of the special theory of relativity. Thus I have presented the method and the most important results of special relativity. It will be appropriate that I give you an exposé about the general r[elativity] th[eory] [ ] in the other separate lectures.

NOTES

1. Typescript has a (superfluous) "relativ" here.
2. Typescript has an (erroneous) "nicht" (not) here.
3. Typescript has an (extra) "relativ" here.
4. Reading typescript's "sagen" as "fragen."
5. Typescript: "trans(formale)versale."
6. Typescript has (mistakenly?): "Elektrizitaets-."
7. Assuming "Auslassung" should be "Auffassung."

8. In German typescript: “Axiome.”
9. Typescript has “transformalen.”
10. Literally: from/of the so-called (“von dem sogenannten”). Maybe “so” should be “vor.”
11. “Bezeichnung” in the typescript should read “Beziehung.”
12. The word “Vorbehalt“ must be a transcriptional error.
13. Assuming “keine” was mistyped as “kleine.”
14. Typescript has “wenn” (if/when).
15. Typescript’s “Ausnahmen” (exceptions) replaced by “Annahmen.”

SECOND LECTURE

10 May 1921

Today I will tell you the principal points about the basis of the general theory of relativity, and will put off the properly mathematical conditions to the last three lectures. I recall what the general problem of the special theory of relativity was. According to the special theory of relativity, all inertial systems shall be on an equal footing for describing Nature's processes. These inertial systems are characterized regarding their relative state of motion by this, that some system K' executes a uniform translational[1] motion relative to a system K. Further, in addition the special theory of relativity is still based on the principle that every light ray found in empty space, viewed from K as well as from K', always has the speed equal to C. Everything else are mere consequences of these basic assumptions. If we regard this special theory of relativity as a whole, then something attracts our attention. The special relativity principle is true, that is, among the inertial systems K, K' etc. we may consider each of them at rest as legitimately as every other. So, concerning the linear, uniform motion, there is no preference of one state of motion over other states of motion. In this respect the purely kinematic relativity of motion, of which we spoke at the beginning, namely that from the standpoint of the concept of motion one can speak only of the motion of one system against another one. This [relativity] is extended also to the physical properties of the motion, namely in the sense that now relativity is not only due to the kinematic properties but that one can also say the states of motion are equivalent in their physical meaning.

What is then remarkable is that this equivalence of states of motion is claimed only for straight-line uniform motion, but not for others. Now this of course urges the question, is it true that the natural laws are such that we do nevertheless have a certain number of preferred states of motion, namely, the states of motion of the inertial system, which are physically distinguished by the particularly simple form assumed by the laws when we avail ourselves of such an inertial system. From a purely philosophical point of view, this concept is rather unsatisfactory, for once the preferred state of motion is not unique[2], in the physical sense[3], one would assume that there isn't any preferred state of motion. But of course this is only an aesthetic motivation, and countering this there is an extraordinarily strong doubt against the possibility of a [suggestion] that the relativity system should not be uniform relativity. This doubt is based on the fundamentals of mechanics. The fundamental statement of mechanics is the principle of inertia, and this states that in relation to an inertial system a point mass left to itself, which does not interact with other masses, moves rectilinearly and uniformly. So in respect to the system K, any particular autonomous mass point would have to move rectilinearly and uniformly. Now one sees at once that this inertial principle cannot apply to definitively moved coordinate systems. For if we now refer to the coordinate system K', and do not say that the coordinate system K' describes a uniform motion, but a rotational motion in the sense of the arrow, then in describing the motion of this mass from K', we will find that the same motion, which is uniform, rectilinear from K, is curvilinear when viewed from K'.[4] So we can say, if we admit arbitrarily[5] conceived systems of motion, then certainly mechanics in its classical form cannot even be upheld in its foundations, but a deep change of mechanics must be needed to remain in agreement with experience. For one would conclude from the failure of the

principle of inertia in the system K' that we may not use the system K', but this conclusion would be incorrect because we do not know whether the principle of inertia is the [whole truth] or if it is only a certain kind of approximation expression as used by physics to date. There is indeed a very important and general fact of experience, which encourages us to believe that a generalization of the relativity principle is quite in conformity with the laws of nature. This general experiential fact is as follows: it is known from geometry that in the gravitational field of the earth at a point on the earth's surface all bodies fall with the same acceleration without regard to what kind of material makes up the bodies. If I drop at the same time a body of wood and one of lead then, apart from air resistance, the law of free fall will be exactly the same for these two bodies. From the [ ] of classical mechanics this fact can be expressed still differently. We speak of two different fundamental properties of matter, namely, first of the inertia of matter and second of the heaviness of matter. The properties of a body in being inert, rather, how inert a body is, is measured by a constant called the inertial mass of the body, that is, a constant belonging to the body, which can be determined by experiments. If we allow a spring stretched in a particular way to act on two different bodies in turn, then they will be accelerated by different amounts, and we say now that the [inverse] value of the acceleration is proportional to the inertial mass. This inertial resistance of the body to acceleration is measured by the inertial mass. This is the same natural constant that we measure by means of a scale. Now there is the law of experience that the inertial mass and the heavy mass of bodies is the same constant, that is, that inertia and heaviness are summarily defined by the same constant, although the definition says that these two properties of the bodies are quite independent of each other. And that is now a very remarkable phenomenon for which classical mechanics has no interpretation. This assumption, inertial mass equals gravitational mass, and the assumption, all bodies fall [with the same acceleration], these two assumptions are equivalent, one can show one is the consequence of the other. This remarkable coincidence immediately loses its puzzling character if we lift ourselves to the view that Nature is constituted such that the relativity principle possesses a still[6] more general meaning. Let us go ahead and imagine a system K as an inertial system in the sense of classical mechanics. We will assume that in reference to the system K, a gravitational field is not present, that is, the physical behavior of a mass shall be such that it remains in the state of rest, if we place it here with ---- if we give it an initial push. Now the question is, how does the mass in respect to the K system relate to K'? We picture a mass that has velocity 0 relative to K. Relative to the system K', which moves with an upward acceleration, this mass will move in accelerated motion downward, that is, the observer from K will say, the mass stays at rest, but the plane of the coordinate system K' moves with upward acceleration, and K' remains at rest and the mass moves uniformly accelerated downward. If we have several masses of different materials, and we refer them to the system K', then these masses will behave mechanically in exactly the same way. The observer, who makes observations from the K system will say, not the mass but the K'[7] system is accelerated, the one from K' will state, these several masses all move with the same acceleration, relative to the system K'. If we did not know a physical space in respect to which there is a law that masses of different character experience the same acceleration, then this statement would not ring a bell.[8] But yes, we can describe our experience, how the motion of the mass proceeds in reference to K', simply by saying, relative to the accelerating system K' the mass behaves exactly as if the system K' were at rest but a gravitational field were present, under the influence of

which the bodies are falling. And it is very suggestive to say, the relativity principle can reasonably be extended to the system K' as well, we must then only recognize that in respect to the system K' there is a gravitational field, and we have no right to say, the system K is [preferred], for such an assumption would exceed the observed matters of fact. We can only certify how the masses behave in respect to the system K', and they behave so [ ] and we can therefore hypothetically state at least for now that the two systems K and K' are entirely equivalent for the physical description, but that a gravitational field exists in reference to K'. This step would be totally physically untenable and fantastical if we did not have in the gravitational field a state of space that has the property of imparting the same acceleration to bodies of different nature. The circumstance that the gravitational field has this property encourages us to interpret this present case in this way, because it is only the re-interpretation of an already physically interpreted case. This hypothesis shall be called equivalence-hypothesis. Once we already see that from the standpoint of physical experience it is suggestive to consider these two states of motion as equivalent, in the sense that one is justified in treating both as rest systems of coordinates; that the presumption about the equivalence of the systems of reference may well be applicable still more generally; and therefore the hypothesis suggests itself that the mutually arbitrarily moving coordinate systems are equally valid.

What is the appeal and the interest that such an investigation can offer? That can be easily appreciated. When we have a coordinate system that executes an arbitrary motion, then relative to this coordinate system every material point will execute some possibly quite complicated motion. If we then treat this coordinate system as a rest system, then we will have to interpret this as the motion of material points in gravitational fields. So in this way we have purely speculatively found out something about the way masses move in gravitational fields and how gravitational fields are constructed architecturally, that is, what are the differential laws of the gravitational field. We can generate certain sorts of gravitational fields artificially, as it were, by introducing, instead of an inertial system, a somehow moving coordinate system, and we can analytically study the laws obeyed by these fields. One always deals here with special cases. When we start with the case of absence of a gravitational field from a geometrical system of reference, we can study certain sorts of gravitational fields, which we can generate by [motions] of the coordinate system, and it then requires an act of generalization to determine the general law. This is not the method one really uses mathematically to find the laws of the gravitational field, but by way of idea this is the method, and the path is indirect. This equivalence principle is pivotal to the general theory of relativity resp. the gr[avitation]al theory and one must have sharp reasoning power.[9]

If this is understood to this point, the execution of this theory would seem to be child's play. It looks as if there are no particular obstacles to progress. One investigates primarily the conditions in the coordinate system, instead of D one has a gravitational field to consider and there we will be able to study the [ ] in our arbitrarily moving coordinate system. But a remarkable and far-reaching difficulty is present, which allows this method appear not straightforwardly applicable, and which demands much deeper reasoning than what seems to follow from this simple reasoning. If we admit coordinate systems in arbitrary relative motion and want to create gravitational fields in this way, it appears that our physical interpretations of space and time fall apart. In yesterday's study of the special theory of relativity, we saw that it is highly significant that the coordinates and the [ ] are

quantities that are results of physical measurements, quantities, quantities that have a very definite physical meaning. Let us consider once more what needs to be presupposed to make this physical interpretation of space and time possible. We first introduce a system of Cartesian coordinates, that is, an inertial system, and with respect to this we determine the coordinates of the several origin[al] points. This construction presupposes the validity of [ ] for the placement of rigid bodies. We speak of this coordinate axis as a straight line. All these concepts presuppose the physical meaning of the concepts of [Euclidean] geometry. Indeed, this Cartesian coordinate system is nothing else but a cubic lattice of rigid cubes[10] and only when we construct this, which fills space up uniformly, can we determine the coordinates in simple fashion by counting the lattice rods. Clocks work in a similar way. About the clocks we postulate that we will not encounter contradictions when we locate them at different places and adjust them with the help of the constancy of light propagation, that all of them will run equally fast. But now simple reasoning shows that all these concepts, upon which we based the space-time quantities as physical quantities, fall apart when we seriously investigate the general theory of relativity. Namely, we see this most conveniently on the example of a system that executes uniform rotation relative to an inertial system.

We will take the blackboard as our inertial system, and we will assume gravitational fields are not present, so an isolated mass moves rectilinearly, uniformly. Light propagates according to the principle of the constancy of the speed of light. ---- This disk rotates on the blackboard in the direction of the arrow. Then the question is, do the laws of Euclidean geometry apply on this disk? On this disk, imagine an observer with equally long rods so he can execute geometrical constructions with his rods, always supposing that these are all equally long segments. There is a question whether these figures, which the man can lay out on his disk, obey the laws of Euclidean geometry in the sense that the rods with which he operates can be considered equal [Euclidean] segments. This is not the case and xx and is proved from the findings of the special theory of relativity. We can say that relative to the blackboard, the special theory of relativity may claim validity throughout. We will imagine the man on the disk[11] laying his sticks along a diameter, and also along the periphery of the disk, all the way around. Now if the laws of Euclidean geometry are valid on the disk, then the ratio of the sticks must ---- equal to the well-known number [π]. But we can easily see that this is not the case on the rotating disk. To see this we have to clarify what this figure is to mean. We think of this figure as an image of the laid-down sticks along with disk circumference at a certain time T, in the inertial system, say, of the blackboard. How must we draw this figure? Here are sticks, those which [move] -- perpendicular to their --. Now we know from the special theory of relativity about the ---- of moving bodies the following: in the transverse direction bodies experience no change in length when in motion, in the longitudinal direction a contraction. So I must draw the sticks on the periphery as shorter in our snapshot than the sticks that are laid on the diameter. If the figure were to be drawn as on a disk at rest relative to the blackboard, then this would not be the case. Then we would, for the ratio of circumference ---- but because the sticks ---- then we will be able to put down a larger number of sticks on the circumference than it would be for the rest case of these figures, and for the ratio of circumference to diameter we get a number, which is larger than π, because for the rotating disk the placing laws of Euclidean geometry are not valid. We can conceive of the rotating disk in the sense of the equivalence principle such that we may also consider the disk at

rest, but then we must assume that a gravitational field is present relative to this disk. As regarded from the blackboard ----

Thus we can interpret the result we found on the rotating disk by saying, if a gravitational field is present, then Euclidean geometry is no longer valid for the placing laws of rigid rods. It is clear that we are no longer able to construct [ ] in respect to this disk, because this construction presupposes the laws of Euclidean geometry for the placing of solid rods as correct. The circumstances are similar for the rates of clocks. For, if we take two equally constructed clocks and put one on the circumference and the other at the center of the rotating disk, then we can again assess the behavior of these clocks from the blackboard[-frame] by applying the insights of the special theory of relativity. Then we find that the clock[12] on the circumference runs more slowly than at the center, because the disk has a [ ] speed, whereas the clock that sits at the center [has] no speed relative to the blackboard. So, viewed from the blackboard this clock runs more slowly than that ----.

From this we will conclude that the existence of a gravitational field has the effect that the clocks located at different places run at different rates, and so we can no longer define time with the help of equally constructed clocks. Thus, we are placed in a difficult position in the general theory of relativity, because we now cannot ascribe a physical sense to the spatial and temporal statements, and we are now really at a loss how to understand space and time at all in a gravitational field site. However, the answer to this question is made much easier by a solution that is already available in mathematics. For a very similar problem as the one that arises here, and arose in surface theory, was treated by Gauss, its formulation as well as its treatment. For Gauss asked himself the following question: some large surface is given, and the geometry of this surface is to be found, that means the following: we imagine small rigid rods, let us say again a very large number of identical rodlets[13], and if we ask about the geometry of the surface. We want to get to know the laws according to which these many, equally constituted rigid rodlets can be placed on this surface. This problem shall now be treated according to a well-defined method, and that is of the essence. The theoretical description of these placement possibilities shall use no geometrical concepts that in any way refer to points in space off the surface; one should stay within the surface. Therefore ask, without using any concepts, which [involve] the space exterior of the [surface]. How can one find out?

When dealing with a plane, it is very easy. There one does it in Euclidean geometry in the following way: one places a Cartesian [grid] and the individual points are determined by numbering the meshes of the grid; 5 meshes to the right and 246 meshes upward and these mesh numbers are the coordinates. In this way, we can describe the positions of the individual points by expressing through these numbers where a point lies. So we can say that the use of Cartesian coordinates is associated with ---- [ ] a physical resp. metrical meaning. Thus, one can find this point with a small measuring rod. But if we want to proceed in a similar way on a large surface it does not work; it fails because we cannot construct such quadratic grids on a large surface. We can say that if we confine ourselves to the surface itself, the positioning laws of Euclidean geometry do not apply on the surface. I want to discuss this in more detail. If we place a small square on the surface, it will work with four small rods and we will [ ]. As long as we move around on a very small part of the surface, everything works fine, but if we cover a larger part, it works less well. The square construction, which leads us to the Cartesian coordinate system, cannot be realized. Here we have the same case as we had in [ ], in the case of a gravitational field.

We are no longer able to define the points in space in a way that allows us to determine the meaning of the coordinates by measurements in a simple way. We can no longer [ ] that elemental construction in metric []. Now Gauss has embarked on the following method. He makes use of this:

1. that for a small piece of such a surface the Euclidean geometry of the plane still plays a role and
2. since we have seen that there is no such metrically meaningful coordinate system on the surface, Gauss proceeded in this way: place quite arbitrary systems of curves on the surface and number the curves of each system.

Now by giving 2 numbers one can naturally [ ] if one specifies how far the curves of this system [ ], so one has assigned[14] numbers to this point. But note, these numbers in and by themselves are not endowed with definite metrical meaning. I cannot simply say, if I start with some point here, and I know the coordinates of this point were 4 and 2, however then this point is impossible to construct because I just do not know how the curves were drawn on the surface. But it turns out that it is nevertheless possible with two such arbitrary systems of curves to describe completely the geometrical property of the surface, and it comes from the following. The geometrical properties of the surface are completely determined if I know for all pairs of neighboring points on the surface how far they are apart on the rodlet scale. If the distance, which I will denote by ds for two neighboring points, if I can express that [distance] in the arbitrarily chosen so-called K system of coordinates, then I can derive all geometrical properties of this surface. From the fact that in small pieces the surface behaves geometrically like a plane, one can conclude that this ds is expressed so that ds^2 is a homogeneous function of second degree of Gauss' coordinate coefficients. This record of information used by Gauss to calculate the geometry of the surface consists in the refusal to give the coordinates an immediate physical resp. metrical meaning, instead to reintroduce this geometrical meaning only indirectly into the theory by expressing the distance between arbitrary points [ ] as a function of the coordinate differentials ---- These functions that then occur are [ ] from case to case, that is, in mathematical expression they are functions of X_1X_2. The analogy between the general theory of relativity and surface theory goes farther, namely, so far that also for infinitesimal smallness there are correspondingly simple metrical laws. In the [surface theory] Euclidean geometry applies [ ] in the general theory of relativity, that is, in an arbitrary space in which there are gravitational fields, that for a small piece of this space the laws of special relativity in fact apply, that means, the coordinates [ ] have direct physical meaning ---- and the metric [??] applies.

So you again see this analogy is a proof of the fact that the laws of the special theory of relativity may really be applied for a small region of the universe.[15] As an example, consider the gravitational field of the earth. We have seen that, in the presence of gravitational fields, as placement laws apply ---- but if no gravitational field is present any longer we may say that Euclidean geometry governs the placement laws of rigid bodies.

The behavior of clocks [is governed by] the laws we have indicated. Imagine a man who falls freely downward in a gravitational field. For him, the gravitational field does not exist. When he lets go of a body[16] this body falls relative to ---- So we can say, for a freely falling coordinate system at any place in the vicinity of Earth, there is no gravitational

force. ---- But if we move far enough from the origin of the coordinate system and drop a body here, it no longer falls in a coordinate direction of the system, but at a slant. We can now [determine] the origin of the coordinate system at certain [ ]. Within this space, we can operate as if no gravitational field were present. On a small surface, there is a limit that is so small that one can operate as if there were a plane instead of the surface ---- As Gauss used an arbitrary continuous numbering of points on the surface by introducing the so-called Gaussian coordinates, so must we similarly introduce an arbitrary numbering of events in the ---- of the phys[ical] [?]. That is, kinem[atical] coordinates in the 4 [dimensional] space of physical events. Just as for two neighboring points on the surface, we can make up a quantity, which we denote as [their] distance, which has direct metrical meaning, so can we do this also in respect to the [ ] of the physical events, and the means are given by the fundamentals of the special theory of relativity. I will only add the remark that [this] analogy is the basis for the possibility to apply straightforwardly the mathematicians' geometrical analysis to the general theory of relativity, and why this is the case I will [now] try to outline. We have seen that if we start from a Galilean coordinate system and install an arbitrarily constructed coordinate system, that then, instead of constant, we get [variable][17] coefficients of the basic metric quantity ds, and that in the physical theory as much as in the surface theory. These variable quantities, which we denote by [ ] in general theory of relativity, these now describe the gravitational field. Namely, it turns out that in the special theory of relativity, these quantities are constant, and that with the appearance of a gravitational field now instead of these ---- so this variable quantity G describes the gravitational field. But secondly it is [ ] that these variable coefficients g, which describe the gravitational field in the 4 [ ]. This [much] as indication about the formal method used by the general[18] theory of relativity. It was developed in its geometrical aspects by Gauss, R[iemann] [?????].

Now that we have more or less seen what kind of concept formation can lead to overcoming the difficulties that we discovered, I want to return again to the physical matter. According to the general theory of relativity, the physical laws must be such that they are valid in 4 dim[ensional] space independent of the choice of the Gaussian coordinate system, and now the question is, how can these laws be found? [What] helps us here [are] just those mathematical theories, which the geometers have established. For example, to discover the laws of the gravitational field, one asks, what kind of equations can be formed from the gravitational pot[entials], what [are] the properties of remaining invariant upon an arbitrary def[ormation] of the coordinate system, can laws be put up which stay unchanged upon each transformation of the coordinates? This condition is a very strong restriction in respect to the physical possibilities and it turns out that if one demands that, namely, as in Newtonian theory, no higher [derivatives] play a role [ ] for this a differential law of gravitation-[?] that then this law is completely determined. Therefore, there is just one possibility for framing differential laws of second order, which must satisfy [ ], and that is the fundamental [reason] why the idea of general relativity leads to a definite theory of gravitation. Obviously, it is here as in the special theory of relativity, that the relativity postulate alone cannot provide a theory, but [ ] exceptions (details) must be made (given). I have compared the special theory of relativity with the [ ]. It turns out that to the relativity ---- only one condition ----

So you see that by this method one is led to certain laws, and the interesting laws to which one is led are the laws of the gravitational field itself, and in conclusion I just want

to tell to what results these investigations have led, to the extent that they are physically interesting. The first thing to be demanded of a gravitational theory is that it lead to Newton's theory as accurately as Newton's is tied to the astronomical experiences over the centuries. Indeed, the outcome is that although the general theory of relativity from a metaphysical standpoint is extraordinarily different from Newton's theory, for it contains ---- and not only a single one, like Newton's theory. In spite of that, it has been shown that for small relative velocities and not exceedingly large masses of the bodies, which move relative to each other, the mechanical laws come to the same as in Newton's theory, with that same accuracy as what is generally reached by astronomy. This is already a big success for the general theory of relativity, for one must bear in mind that Newton's theory contains a arbitrary assumption similar to the assumption concerning the law of how[19] the heavenly bodies act on each other. ---- The general theory of relativity, in contrast, yields the law of attraction from purely formal arguments, without the need to introduce such an arbitrary hypothesis. But upon somewhat more accurate calculation, the general theory of relativity yields certain deviations from Newton's laws, in particular, first related to the motion of the heavenly bodies[/force?]. Namely, according to Newton's laws a planet describes an ellipse, which is fixed relative to the fixed stars ---- Here the orbit would [ ] ellipse of a planet, which [ ] in this ---- This major axis shall according to ---- But if one calculates exactly according to the general theory of relativity, one finds a very slow rotation of this ellipse relative to the fixed stars, and this very slow rotation for the sun's closest planet Mercury has been known for a century. ---- In the same sense, in which [ ] of the planets takes place, for Mercury amounts to about 42 seconds of arc per century. This amount also comes out exactly like that from the general theory of relativity. At the same time, one sees that this [precession] then decreases with the mean radius of the planet['s orbit] so quickly that a verification [is] not possible for the other planets. This is a very remarkable success of the general theory of relativity.

The second consequence, which has also been possible to test already, is that of the deflection of light rays. This has caused particular sensation because it was derived before the experimental fact was known. The theory predicts that a light ray passing the sun experiences bending by the gravitational field of the sun, a bending by about 1.5 seconds of arc. That this curving of light rays by the gravitational field of the sun in fact takes place was confirmed by the two English solar experts, who exposed [ ] on 29 May [in] Brazil and on the West coast of Africa. During the eclipse, stars near to the sun ---- Let this be the sun, and here shall be the earth, and we will now say in this direction, which I am drawing, there shall be a fixed star, practically at infinity. If the sun is in place here, then by the theory this light ray shall experience a deflection and therefore does not get to the observer. To see in which direction the observer sees the star, I must therefore take a light ray that gets to the observer in spite of the deflection. This small angle here, which adjusts itself between the immediate direction to the star and the direction in which the star is seen, can be determined by experiment. The simplest [way], and also the way it was actually done[20] [is] to consider two stars, which are located on different sides of the sun ---- This is the true direction in which the star is located, and this [is] that line, which [follows] the path of the light ray under the influence of the deflecting[21] force of the sun ---- Then one sees the following: if the sun is not there, this angle here is the angle that apparently obtains between the two stars, because then the light rays proceed rectilinearly, but when the sun is there, then the angle that is noted between the two stars is larger, and by

fixing photographically and picture-taking under otherwise equal conditions [ ] one can [ ] the [ ] force of the sun. One must say that the English experts did a good job, for they have shown a positive result and confirmed the prediction of the theory.

Now there is still a third consequence, namely, that of the theory, which is not yet definitely settled concerning the motion. Namely, from the example of the rotating disk, we saw that a clock located at [ ] than the clock at the center, and from that we concluded that clocks at different places run at different speeds. Thus, if one clock is located at the sun's surface and another [one]—identical clocks—then these clocks would run at different speeds, to be exact, at the sun's surface by about [1/(]2 million[)] more slowly than at the earth's surface. Nature has given us such objects, which perform with a di[fferent] temporal rate, and can be considered as models of clocks. We know that a central line ---- It has an inner radius ? and it is now to be concluded from the theory that those entities that emit in the sun, emit more slowly when viewed from earth and therefore the spectral lines shall be shifted toward the red compared to the corresponding terrestrial lines. The astronomers are debating whether this relativistic shift of spectral lines exists or not. From the standpoint of the general theory of relativity one can say, if it doesn't exist[22] then the theory is refuted. Now, in a short discussion, I want to go into the general significance of the general theory of relativity. The generally interesting innovation brought on by the general theory of relativity is in essence the very essential change in the conceptualization of space and time compared to the special theory of relativity. Namely, when we speak of space, we mean by that nothing other than the laws that dictate the placing of bodies. No different meaning can be given to statements about space. Now we saw that the laws for positioning light [rays] depend on the gravitational field. The gravitational field itself depends in turn on the distribution of bodies. So we come to the statement that the laws of placing bodies depend on the position of the masses, and therefore the laws of spatial placing are endowed with independent existence, exactly like the bodies. So it seems that a judgment about the spatial and the temporal is so intimately connected with the remaining physical laws that it can in no way be separated from them. We can no longer proceed by first establishing space and time when constructing the realm of physics, and only later the objects that move to some extent in this container. Rather, the reality is that the spatial and the temporal are unmistakably connected with [everything] else, and with this I have finished my general considerations and I will begin to treat the matter from a more mathematical standpoint.

NOTES

1. Typescript: transformational.
2. Literally: once there is already no one preferred state of motion.
3. "Bezeichnung" (sense, notation) reinterpreted as "Beziehung" (sense, respect).
4. Typescript has "K."
5. Misreading of "bildlich" of typescript for "beliebig."
6. Assuming "nicht" (no) of typescript should be "noch" (still, even).
7. Typescript has "K" here.
8. Literally: would not make us see the light ("wuerde uns . . . kein Licht aufgehen lassen").
9. Typescript has "Ueberlegenheit" (superiority); "ueberlegen" can mean "to reason" or "to be superior."
10. Typescript has misunderstood "Guertel" (belts) for "Wuerfeln" (cubes, dice).
11. Typescript: "auf der Tafel" (on the blackboard).
12. Typescript: "Scheibe" (disk).

13. The German "Federstab" (spring rod, torsion bar; feather teaser) is in puzzling contradiction to the rods' rigidity.
14. Assuming "so geordnet" should be "zugeordnet."
15. Literally: a small part of the world ("einen kleinen Teil der Welt").
16. Literally: if a body makes itself loose ("Wenn sich ein Koerper losloest").
17. Typescript: "sterio."
18. Typescript: "spezielle" (special).
19. German: "welcher." Unless this is typo, grammatically the antecedent must be assumption ("Annahme"), not law ("Gesetz").
20. Typescript: "geschah" (happened).
21. Typescript "abblitzend" (rebuffing).
22. Literally: if it's not in tune ("wenn es nicht stimmt").

PART V

BIOGRAPHICAL NOTES ON THE PROTAGONISTS OF THE FORMATIVE YEARS

BIOGRAPHICAL NOTES

This collection of biographical notes may be consulted selectively from the text or read sequentially to get an impression of the scope of scientific networking during the formative period. To help those who prefer the second option, we have added at the end of each note the context in which an individual is mentioned in the text.

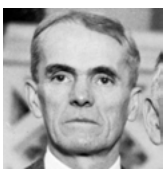
During its formative years, general relativity changed from the esoteric pursuit of one individual corresponding with a handful of colleagues into a topic of interest of a growing and highly engaged community of scientists and philosophers. The following section of short biographical notes aims to illustrate this phenomenon. These notes are not confined to simple biographical information. Rather, they are intended to reflect involvement in specific points raised in the book. Some of the individuals mentioned therein, those who played a more peripheral role in its narrative, are not included in this section.

Abraham, Max (1875–1922)

Photo: SUB Göttingen, Sammlung Voit : M. Abraham, Nr. 1.

Max Abraham, a German theoretical physicist, completed his studies in Berlin under Max Planck, after which he served as Planck's assistant for three years. He is known for his work on electrodynamics and electron theory. In 1902, he developed a theory of electrodynamics in which he applied Maxwell's differential equations of the electromagnetic field to the dynamics of electrons. In 1912, he formulated a theory of gravitation using Minkowski's four-dimensional spacetime formalism. Abraham was one of the first to address the issue of gravitational waves in comparison with electromagnetic waves and concluded that such waves can hardly be measured. In the early years of Einstein's work on general relativity, he became one of his main opponents. (pp. 94–95, 96, 100, 102)

Adams, Walter Sydney (1876–1956)



The American astronomer Walter S. Adams studied at Dartmouth College, the University of Chicago, and the University of Munich. He joined the Mount Wilson Observatory as a regular member of its staff in 1904 and served as its director between 1923 and 1946. He worked on stellar spectroscopy and investigated sunspots, the rotation of the sun, and the relation between spectral line intensities and the absolute magnitude of a star. His studies enabled him to determine distances and velocities of thousands of stars. In 1915, he began to work on the smaller and massive companion of Sirius (Sirius B), making history by discovering the first white dwarf star. Based on extended measurements of the spectrum of Sirius B (between 1915 and 1925), he claimed to have

confirmed the gravitational redshift predicted by Einstein's general theory of relativity. Adams played an important role in the planning the 200-inch telescope for the Mount Palomar Observatory. He hosted Einstein during his visit to Pasadena. (pp. 61, 64–66)

Besso, Michele (1873–1955)

Photo: Courtesy the Besso family, Lausanne, Schweiz

Einstein met Michele Besso, a mechanical engineer, in 1896 in Zurich, where he had recently enrolled as a student at the Federal Technical University. They became lifelong friends and, for a while, were even colleagues at the Swiss Patent Office in Bern. Besso was an interested listener who was able to pose questions that would stimulate Einstein, in their frequent and long conversations, to discuss, clarify, and develop his ideas. Besso's role in the genesis of special relativity was so important that he was the only person to be acknowledged by Einstein in his paper. Following the publication of the *Entwurf* theory, they collaborated on the calculation of the Mercury perihelion motion and of the gravitational field in the interior of a rotating mass shell to demonstrate the relativity of rotational motion. Also during the formative years, Einstein corresponded regularly with Besso on his scientific agenda. (pp. 53, 61, 75, 80, 86, 95, 106, 142)

Born, Max (1882–1970)

Photo: Archiv der Max-Planck-Gesellschaft, Berlin-Dahlem

Max Born was one of the founders of modern quantum physics. He studied mathematics and physics at different universities. Among his teachers at the University of Göttingen was Hermann Minkowski, who introduced him to electrodynamics and special relativity theory. Born's first publications, written between 1909 and 1914, were dedicated to electron theory, relativity theory, crystal physics, and Einstein's quantum theory of specific heat. In the following years, he focused his work on atomic physics and the mathematical development of quantum physics. In 1915, Born was appointed professor for theoretical physics at the University of Berlin where he became a close friend of Einstein. He later went to Göttingen, where he formed a group that in 1925 formulated the foundations of quantum mechanics. Born worked on special relativity but nevertheless made a significant contribution to the understanding of general relativity with his popular book published in 1920. (pp. 9, 94, 102, 149)

Campbell, William Wallace (1862–1938)

Photo: Courtesy AIP Emilio Segre Visual Archives

William Campbell was an American astronomer. He began his academic career as a professor of mathematics at the University of Colorado but moved back to his alma mater, the University of Michigan, as an instructor in astronomy. Beginning in 1981, he worked at the Lick Observatory in California, where he became a pioneer in astronomical spectroscopy, cataloguing the radial velocities of stars. Between 1900 and 1930, he served as the director of the observatory, and from 1923 to 1930, as president of the University of California. In 1922, he led an astronomical expedition to the North West Coast of Australia, which reconfirmed the 1919 Eddington expedition's observations of

the bending of light by the gravitational field of the sun as predicted by Einstein's general theory of relativity. (pp. 57, 65)

Cassirer, Ernst (1874–1945)

Photo: Center for Advanced Research in Phenomenology–CARP

Ernst Cassirer was one of the leading German philosophers of the first half of the twentieth century. In Berlin, Marburg, and other cities, he studied law, German literature, and philosophy. He became a member of the so-called Marburg school of neo-Kantianism, which focused on epistemology in the Kantian tradition, attempting to take into account the results of the natural sciences. In 1906, he habilitated in Berlin with a major study on epistemology and its history, comprising both the history of philosophy and the history of science. In his study on the concepts of substance and function, he emphasized the relational character of the concepts of the sciences and the symbolic function of the language of mathematics. From this perspective, he developed an interest in relativity theory, which he saw as a realization of the modern tendency to replace concepts of substance by concepts of functional relations. Although Einstein was skeptical about Kantian philosophy, he appreciated Cassirer's work. Using the concept of symbolic forms, Cassirer subsequently extended his studies to a philosophy of modernity centered on the role of active interventions in the world and that of the representation of its perceptions. Between 1919 and 1933, Cassirer was professor of philosophy at the newly founded Hamburg University, where he became the first Jewish rector of a German university; in 1933, he was forced to leave Germany. (pp. 8, 9, 113, 146)

Droste, Johannes (1886–1963)

Johannes Droste received his PhD at Leiden for his work under Hendrik Lorentz on the gravitational field of one or more bodies according to Einstein's theory. Initially, he used as starting point the Einstein-Grossmann *Entwurf* theory. In 1916, he wrote the paper "The Field of a Single Center in Einstein's Theory of Gravitation and the Motion of a Particle in That Field," which was already based on the generally covariant field equation. In this paper, he independently produced the same solution as Schwarzschild but using a simpler, more direct derivation and representation. Droste's contributions were applied by de Sitter to astronomical problems. These papers appeared only in Dutch and went largely unnoticed. Some of their results were, unknowingly, rederived by Einstein, Infeld, and Hoffmann in their work on the problem of motion in general relativity in the 1930s. (pp. 8, 47, 49)

Duhem, Pierre Maurice Marie (1861–1916)

Pierre Duhem studied at the École Normale Supérieure in Paris. He later taught at the Universities of Lille, Rennes, and Bordeaux. In 1913, he became a corresponding member of the French Academy of Science. He was one of the founders of physical chemistry and is best known for his work on chemical thermodynamics. He also made major contributions to the philosophy and history of science. His views were inspired by his Catholic and anti-republican religious and political stances. He tried to show that important roots of modern science can be found in medieval philosophy and,

in this way, opened up new views to the history of science. In the philosophy of science, he argued that a scientific theory is not completely determined by observational data and advocated a holistic view of scientific theories. In 1915, he published an essay polemicizing against relativistic physics as a characteristic expression of German science, confounding sound common sense. (p. 114)

Eddington, Arthur Stanley (1882–1944)

Photo: George Grantham Bain Collection, Library of Congress Prints and Photographs Division Washington, D.C.

The British astrophysicist Arthur Eddington became director of the Observatory of Cambridge in 1914. His research focused on stellar physical processes such as radiation and energy generation and later on the mathematical and cosmological aspects of general relativity. Eddington became aware of Einstein's general theory of relativity in 1915 through Willem de Sitter's papers and soon began to promote its empirical testing. In 1919, he led an expedition to the African island of Príncipe to observe the behavior of light rays in the gravitational field of the sun during a solar eclipse. Eddington's observations confirmed the deflection of light as predicted by Einstein's theory and were considered at the time as conclusive proof of general relativity. During the formative years of general relativity, Eddington corresponded extensively with Einstein on the cosmological consequences of the theory and on his ideas about the unification of the gravitational and electromagnetic fields. (pp. 8–9, 48, 56–57, 60, 62, 72, 101–102, 117, 125, 127–130, 144–145, 155, 404)

Ehrenfest, Paul (1880–1933)

In 1912, Paul Ehrenfest was appointed to the chair of theoretical physics at Leiden in succession to Hendrik Antoon Lorentz. In the preceding years, he and his wife Tatiana Afanasieva made important contributions to statistical mechanics. In 1909, Ehrenfest formulated a paradox showing that rigid bodies are not compatible with the special theory of relativity. This led Einstein to the recognition that the geometry of a rotating frame of reference is not Euclidean. Ehrenfest worked predominantly on the early quantum theory and later on quantum mechanics. After meeting for the first time in 1912, Einstein and Ehrenfest had many occasions to discuss various physics problems; in particular, Ehrenfest was an active participant in the discussions of the cosmological consequences of general relativity. Ehrenfest also arranged a number of notable conversations between Einstein and Niels Bohr on quantum physics. (pp. 11, 47, 59, 70–71, 108, 127)

Freundlich, Erwin Finlay (1885–1964)

Photo: Archiv Juergen Staude, photograph by Erich Strohbusch

Erwin Freundlich completed his studies in mathematics at the University of Göttingen. At the suggestion of his tutor, Felix Klein, he joined the observatory in Berlin, where he became the first astronomer to work in association with Einstein to test predictions of the general theory of relativity by astronomical observations. Freundlich enthusiastically took on the task of testing the effects of gravitation on the propagation

of light, as predicted by the still-incomplete theory of general relativity. For a while, he was one of Einstein's closest collaborators but remained isolated from other German astronomers because of their skepticism toward Einstein's theory. In 1920, he was the principal initiator of the Einstein Institute, an observatory dedicated specifically to strengthen the empirical foundations of Einstein's gravitational theory in the astrophysical domain. During his scientific career, he organized several expeditions to observe light deflection during solar eclipses and focused his research on the wavelength shifts in solar and stellar spectra. With the rise of Hitler to power in 1933, Freundlich had to leave Germany because of his Jewish heritage. (pp. 8, 54–55, 57–58, 59–60, 111, 141, 404)

Friedmann, Alexander Alexandrovich (1988–1925)

The Russian physicist and mathematician Alexander Friedmann is best known for his pioneering solutions of Einstein's gravitational field equations, which indicated the possibility of an expanding universe at a time when Einstein was still strongly committed to the idea of a static universe. His analysis preceded that of the Belgian cosmologist George Lemaître. In a paper published in 1924 in the *Zeitschrift der Physik*, he presented three possible scenarios of the evolution of the universe (expanding, static, and contracting), known as the three Friedmann models describing positive, zero, and negative curvature, respectively. Through this work and his travels through Europe in 1923–1924, he became part of the founding group of the emerging field of relativistic cosmology. His premature death from typhoid fever at the age of 37 cut short a career that would have undoubtedly continued to impact the evolution of modern cosmology. (pp. 8, 50, 82–83, 85, 86, 90, 272 [112], 287 [127], 395, 398)

Grommer, Jakob (1879–1933)

Jakob Grommer, a Russian, Orthodox Jew, went to the University of Göttingen in 1912 to study mathematics. He had had no previous education, except in a Talmudic school in Brest-Litovsk (today's Belarus). Hilbert quickly recognized Grommer's outstanding talent and provided him with the necessary support to complete a doctoral degree in 1914. With the disadvantage of his Jewish-Russian origins and handicapped by a physical deformation caused by acromegaly, a disease that leads to enlargement of the bones, face, and jaw, Grommer could not obtain an academic position. Einstein met him during a visit to Göttingen, and Grommer became his assistant in the years 1916 to 1928, supported by the Kaiser Wilhelm Institute for Physics, of which Einstein was director. In 1928, he obtained an academic position in Minsk. (pp. 70–71, 131)

Hale, George Ellery (1868–1938)

At a time when astronomers were mostly concerned with the positions, motions, and distances of stars rather than their physical nature, the American astronomer George Ellery Hale was already applying spectroscopic methods to the observation of solar phenomena during his student years at MIT and at the Harvard College Observatory. Hale made great contributions to the advancement of astrophysical research

with his observations, in particular of solar spectra and the magnetic fields in sunspots, and by promoting the foundation of astrophysical observatories. In 1904, he founded the Mount Wilson Observatory and directed it until his retirement. He was the key figure in the planning or construction of several leading telescopes at Yerkes Observatory, Mount Wilson Observatory, and Mount Palomar Observatory. It was Hale's authority in solar astrophysics that probably led Einstein in 1913 to seek his advice on the possibility of detecting light deflection in the sun's gravitational field. (pp. 55, 395, 405)

Heckmann, Otto Hermann Leopold (1901–1983)

Photo: Hamburger Sternwarte

Otto Heckmann was a German observational astronomer and theoretical astrophysicist whose scientific research covered astrometry (the positions and proper motions of stars), photometry of stellar clusters, stellar dynamics, and theoretical cosmology. In the 1920s, with Einstein's general relativity becoming the dominant theory of gravitation, Heckmann's interest focused on cosmological models of this theory. His work triggered the Einstein–de Sitter model of the universe. He worked at the observatories of the Universities of Bonn and Göttingen, and from 1941 to 1962 was director of the Hamburg Observatory and head of the Department of Astronomy at Hamburg University. After that, he was one of the founding members of the European Southern Observatory (ESO) in Chile and left a particularly strong imprint as manager for the astronomical community in Europe. (p. 89)

Hilbert, David (1862–1943)

From 1895, David Hilbert was professor of mathematics at the University of Göttingen. He was one of the most influential mathematicians of his time and provided new and enduring impulses in many branches of his discipline, including the theory of invariants, algebraic number theory, analysis and the theory of integral equations. Pursuing a program of the axiomatization of all mathematical sciences, in 1912 Hilbert started to work on the axiomatic foundations of physics. He learned about Einstein's efforts toward general relativity and derived independently and about at the same time a variational formalism from which the gravitational field equations could be derived. Hilbert's interest in the theory of general relativity continued in the formative years, when he re-derived Schwarzschild's solution and made an (not quite successful) attempt to define the singularities of solutions of the field equations. (pp. 4, 9, 25, 29, 47, 48, 123, 396, 401, 407)

Hubble, Edwin (1889–1953)

Photo: Reproduced by the permission of The Huntington Library, San Marino, California

In 1919, the American astronomer Edwin Hubble began working at the Mount Wilson Observatory in California. With his investigations on galactic nebulae and variable stars, he played a crucial role in establishing modern extragalactic astronomy. In a paper published in 1929, Hubble delivered observational evidence that the recessional velocity of a galaxy increases with its distance from the observer. Initially, he was

uncommitted to the conclusion of an expanding universe as expounded some years earlier by Georges Lemaître and Alexander Friedmann on the basis of calculations in the frame of Einstein's general relativity. Hubble's results convinced Einstein to abandon the cosmological constant he had introduced in his field equations in order to maintain the model of a static universe. (pp. 8, 63–65, 84–88, 89–91, 398, 405)

Humason, Milton Lasell (1891–1972)

Photo: Courtesy AIP Emilio Segre Visual Archives, Brittle Books Collection

Humason worked as a janitor at the Mount Wilson observatory and volunteered to be a night assistant to help with astronomical observations. His technical skill attracted the attention of the director, George Hale, who appointed him as staff member in 1917, although Humason had no doctoral degree, not even a high school certificate. Humason became an independent and careful observer, obtaining photographs and difficult spectrograms of faint galaxies and making several important discoveries, including Comet C/1961 R1, which is characterized by its large perihelion distance. His observations played a major role in the development of cosmology. He worked for many years with Hubble, actively participating in the discovery of Hubble's law on the relation between the distance and recession velocity of distant galaxies. (pp. 63, 65, 85)

Kaluza, Theodor Franz Eduard (1885–1954)

Photo: Archives of P. Roquette, Heidelberg and the archives of the Mathematisches Forschungsinstitut Oberwolfach

Theodor Kaluza started his academic career as a mathematician, obtaining his doctoral degree at the University of Königsberg. He was attracted to physics by studying general relativity. In 1919, he extended Einstein's gravitational field equation to five dimensions and showed that this formalism provides a possible scheme for a unification of the gravitational and electromagnetic field. Einstein encouraged him to publish his highly original ideas, which he did in 1921. Kaluza's model is known as the Kaluza-Klein theory, to mark the contribution of Oskar Klein, who characterized the fifth dimension as a small, compactified dimension. The theory, initially a popular topic of research, quickly lost favor with the introduction of quantum mechanics and only reemerged in the context of string theory. In 1929, Kaluza was appointed as a professor in Kiel. In 1935, he was made a full professor at the University of Göttingen, where he remained until his death. The latter appointment was obtained with great difficulty, because he refused to accept the National Socialist ideology. (pp. 9, 125, 130–132, 258 [98])

Klein, Felix Christian (1849–1925)

Photo: Universitätsarchiv Leipzig

Felix Klein was a German mathematician known mainly for his work on group theory, non-Euclidean geometry, and the connections between geometry and group theory. Working at the University of Erlangen-Nuremberg, in 1872 he formulated his highly influential Erlangen program, according to which geometries are classified by their underlying symmetries. After that he worked at the Technical University of Munich

and the University of Leipzig. In 1886, he accepted a chair at the University of Göttingen, where he established one of the best mathematical research centers in the world. He brought David Hilbert from Königsberg to join his research team at Göttingen in 1895. He taught a wide variety of courses, mainly on the interface between mathematics and physics. Klein had a peculiar way of considering the relation between mathematics and physics as a result of his Erlangen program and geometrical modeling of physical phenomena. He was aiming at a physical foundation for non-Euclidean geometry. In 1917, he proposed his work on projective geometry to Einstein to be applied in further developments of general relativity. In those years he also corresponded with Einstein about the ongoing debate around the cosmological consequences of the gravitational field equation and about the meaning of singularities in their solutions. (pp. 8, 9, 29, 77–78, 392)

Klein, Oskar Benjamin (1894–1977)

The Swedish physicist Oskar Klein worked for a few years with Niels Bohr at the University of Copenhagen and received his doctoral degree at the University College of Stockholm in 1921. He then worked at the Universities of Michigan and Copenhagen and Leiden and Lund Universities before accepting, in 1930, a professorial chair in physics at Stockholm University College. In 1924, while still in Michigan, he taught a course on electromagnetism and lectured on electric particles in a combined gravitational and electromagnetic field. This was the beginning of his landmark work on a unified field theory. Klein was apparently unaware of Kaluza's work, and like Kaluza before him, he extended the treatment to five dimensions. However, he proposed a novel interpretation of the fifth dimension, which was supposed to be curled up into a small ball of the dimensions of the Planck length, 10^{-33} cm, divided by the fine structure constant. Thus, the fifth coordinate was not observable, but it constituted a physical quantity representing the electric charge. Despite a lot of initial interest in unification, the Kaluza-Klein theory was abandoned by most physicists and had its renaissance only in the context of present-day string theory. (pp. 9, 131–132, 394, 395)

Kopff, August (1882–1960)

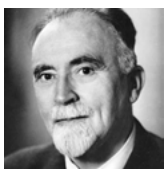


Photo: Astronomische Rechen-Institut (ARI), University of Heidelberg

The eminent German astronomer August Kopff began his career at the Heidelberg Observatory. In 1922, he participated in the solar eclipse expedition to Christmas Island, in the Indian Ocean, to measure the bending of light. Between 1924 and 1954, he was the director of the Astronomisches Rechen-Institut in Berlin and later in Heidelberg, working mainly on astrometry and preparing fundamental catalogues of astronomical data. In addition to astronomical observation, he worked also on the theory of relativity. Between 1919 and 1923, he gave lectures on this subject that were summarized into a textbook. In 1930, Kopff exchanged letters with Einstein on the effect of tidal forces on the evolution of the planetary and the earth-moon systems. (p. 149)

Kretschmann, Erich Justus (1887–1973)

Photo: Universitätsarchiv Halle-Wittenberg, Rep. 40/I, K 62

The German physicist Erich Kretschmann studied under Max Planck and Heinrich Rubens at the University of Berlin, from which he received his PhD in 1914. He became a high school teacher and later a professor of theoretical physics at the Universities of Königsberg and Halle-Wittenberge. In 1917, he published a paper on Einstein's theory of general relativity in which he argued that the principle of general covariance has no physical content but only constitutes a mathematical requirement. This led to an exchange of letters with Einstein and to Einstein's reformulation of the principles of his theory of general relativity. (pp. 8, 9, 37–41)

Lanczos, Cornelius (1893–1974)

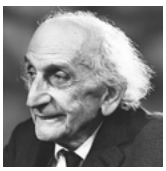


Photo: With kind permission of University of Manchester, Institute of Science and Technology (UMIST)

Lanczos was a Hungarian mathematician and physicist. His original name was Kornél Löwy. At the University of Budapest, he studied with Lóránd Eötvös, who first aroused his interest in relativity theory. In 1921, he completed his PhD on a mathematical aspect of relativity and dedicated his work to Einstein, who appreciated its scientific merits. After completing his PhD, Lanczos left Hungary because of political unrest and anti-Semitism and took up a position at the University of Frankfurt. In 1922, he published a study on the role of coordinate systems in general relativity. Contradicting a widely held belief, he showed that no singularity occurs at the Schwarzschild radius and that one cannot infer an actual singularity of the field from the singular behavior of the metric. It took many years before this insight was generally recognized. He also worked on quantum theory, unified field theory, and cosmology. In 1924, he found an exact solution of Einstein's field equations representing a cylindrically symmetric rigidly rotating configuration of dust. Between 1928 and 1929, Lanczos worked with Einstein in Berlin. In 1931, he spent a year as a visiting professor at Purdue University in Indiana. After his return to Germany, in view of rising anti-Semitism and the ascent of the Nazi regime, he decided to emigrate. In 1932, he returned to a professorship at Purdue where he worked, among other subjects, on applied and numerical mathematics. In reaction to the repressive atmosphere of the McCarthy era, he left the US in 1952 to take up a position at the Dublin Institute for Advanced Studies as successor to Erwin Schrödinger. His 1949 book *The Variational Principles of Mechanics* has become a classic. (pp. 48, 49)

Laue, Max von (1879–1960)

The German theoretical physicist Max von Laue was appointed as a lecturer at the University of Berlin in 1906. From 1914 to 1919, he was a professor at the University of Frankfurt, and from 1919, at the University of Berlin. Von Laue was particularly concerned with mathematical aspects of problems of optics and with theories of X-ray interference with matter. He is best known for his discovery of the diffraction of

X-rays by crystals. In 1907, he delivered a mathematical explanation of light propagation in the framework of Einstein's special theory of relativity. Von Laue's work fostered the acceptance of the theory. Although he was a little slow in accepting the theory of general relativity, he made a major contribution to its development by his work on relativistic continuum dynamics described in his book that was published in 1911. This work led to the energy-momentum tensor, which plays the crucial role of the source term in Einstein's equation for the gravitational field. Von Laue and Einstein first met in 1906 and became lifelong friends. (pp. 9, 28, 48, 49, 117, 145–146)

Lemaître, Georges (1894–1966)

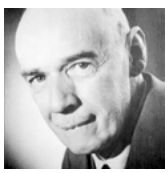
Alongside theological studies, the Belgian Catholic priest Georges Lemaître pursued research in astrophysics, cosmology, and mathematics. In 1927, he was appointed professor of physics at the University of Louvain. As early as 1925, Lemaître worked on the application of Einstein's general theory of relativity to cosmology and in 1927 published a fundamental paper in which he solved the equations of the gravitational field without using Einstein's cosmological constant. He sustained the view that the universe is expanding, as Alexander Friedmann had considered some years earlier. Lemaître delivered a demonstration of Hubble's law of recessional velocity before Hubble himself. He was the first to propose an early version of the Big Bang theory, suggesting that the universe evolved from a primeval atom. (pp. 8, 48–50, 62–63, 66, 83–85, 90, 395)

Lenard, Philipp (1862–1947)

Photo: German Federal Archives

The German experimental physicist Philipp Lenard spent most of his active years at the University of Heidelberg (1907–1931), where he served as head of the Physics Institute and led the Radiological Institute, which he founded. He received a Nobel Prize in 1905 for his work on cathode rays. Lenard also worked on phosphorescence and on the conductivity of metals, but he is best known for his experiments on the photoelectric effect, for which Albert Einstein provided a theoretical interpretation in 1905. From the early 1920s, Lenard became the most ardent opponent of Einstein's theories of relativity and of Einstein himself. This attitude was based on strong nationalistic ideology and on overt anti-Semitism. In 1936–1937, Lenard published *Deutsche Physik* (German physics), which was written on the basis of classical experimental physics of the nineteenth century and rejected theoretical physics almost completely, which he denounced as "Jewish" physics. (p. 152)

Lense, Josef (1890–1985)



Josef Lense studied astronomy and mathematics and later started his academic career as a lecturer at the University of Vienna. From 1927 until 1946, he served as professor of mathematics and later as director of the Institute of Mathematics at the Technical University of Munich. His colleague in Vienna, the physicist Hans Thirring, being aware of his background in astronomy, suggested that Lense use perturbation theory

to derive from Einstein's gravitational field equation the equation of motion of a particle in a field generated by a rotating mass, with application to the orbits of planets and the moon. This resulted in a joint paper and in the discovery of the Lense-Thirring frame-dragging effect. (pp. 80–81, 148, 406)

Levi–Civita, Tullio (1873–1941)

The Italian mathematician Tullio Levi-Civita taught at the Universities of Pavia, Padua and, finally, Rome. He published a great number of works on pure and applied mathematics, dealing in particular with analytical mechanics, celestial mechanics, hydrodynamics, elasticity, electromagnetism and atomic physics. Around 1899–1900, Levi-Civita and his professor, Ricci-Curbastro, wrote a fundamental treatise on absolute differential calculus and its applications for expressing geometric and physical laws in Euclidean and non-Euclidean spaces. Einstein and Grossmann later used and developed these new mathematical tools in order to formulate the general theory of relativity. In the years 1915–1917, Einstein and Levi-Civita corresponded about mathematical aspects of the theory. At that time, Levi-Civita refined the mathematical concepts and methods of Riemannian geometry, which were later adopted into physics by Hermann Weyl and used by Einstein in his formulation of general relativity in 1921. (pp. 19–20, 47, 126, 129, 144, 224 [64], 302 [142], 334 [141])

Lorentz, Hendrik Antoon (1853–1928)

The Dutch physicist Hendrik Antoon Lorentz was appointed professor at the University of Leiden in 1877. His most influential contributions concerned the theory of light and electromagnetism as well as the electron theory. Starting from Maxwell's electromagnetic theory of light, Lorentz developed a unifying theory of the electron and electrodynamics of moving bodies based on the assumption of a stationary ether. Lorentz's electrodynamics constituted the basis of Einstein's special theory of relativity, which, however, disposed of the ether. Lorentz recognized the consistency of Einstein's theory and made contributions to it but continued to support the existence of the ether. In later years, Lorentz also contributed to the development of the general theory of relativity. In extensive correspondence, Einstein shared with him results and dilemmas on his road to the final version of the gravitational field equation. (pp. 4, 9, 12, 19, 28, 29, 39, 47, 106, 391, 392, 401)

Mach, Ernst (1839–1916)

The Austrian physicist and philosopher Ernst Mach was appointed professor of physics at the University of Prague in 1867 and professor of the history and philosophy of science at the University of Vienna in 1895. His physical research was dedicated to problems of optics (the Doppler effect) and acoustics (sound waves). In this connection, Mach also studied the physiology and psychology of sensory perception, on which he published a book in 1886. These studies led Mach to question the mechanistic and atomistic views prevalent at the time and to formulate an empiricist

theory of knowledge that strongly influenced the logical positivism of the twentieth century. In particular, Mach's criticism of Newton's concept of absolute space, which was expounded in a book he published in 1883 on the history of mechanics, played a crucial role in the heuristics that led Einstein to his general theory of relativity and in his cosmological ideas. (pp. 8, 22, 34, 35, 36, 41, 55, 69, 78–79, 81–82, 106, 110, 113, 115, 116, 118, 403)

Mayer, Walther (1887–1948)

The Austrian mathematician Walther Mayer received his doctoral degree from the University of Vienna. Being Jewish, he had great difficulties in obtaining an academic position in Austria. With Einstein's help, however, he was offered a lecturer's position at the University of Vienna in 1926. In 1929, he became Einstein's assistant. They collaborated and published joint papers on unified field theories. Mayer was nicknamed "Einstein's calculator." With the rise of Hitler to power, he followed Einstein to the United States and continued to work with him at the Institute for Advanced Study in Princeton, where he became an associate in mathematics. (p. 62, 64)

Meyerson, Émile (1859–1933)

Photo: The National Library of Israel

Émile Meyerson was born in Lublin and studied chemistry in Heidelberg, Berlin, and Paris. Between 1900 and 1923, he was the director of the Jewish Colonization Association, which supported the emigration of Jews from Russia, especially to Palestine. At the same time, he was interested in the philosophy of science. In 1907, he completed his first book, *Identity and Reality*, followed by a number of further studies that considered the history of science and ideas as a search for rationality in the multitude of phenomena driven by a quest for unity of understanding. He rejected the purely descriptive approach of the positivists, expressing his belief in the possibility of scientific explanation. Einstein corresponded with Meyerson about the interpretation of relativity theory and appreciated his approach. (p. 118)

Mie, Gustav (1869–1957)

Photo: Lepoldina Archiv/M1 3412

The German physicist Gustav Mie was appointed professor at the University of Greifswald in 1902 and later taught at the Universities of Halle and Freiburg. Besides his contributions concerning optical phenomena in colloids, X-ray analyses of organic compounds, and electromagnetism, Mie's most influential work was a theory of matter published in 1912. As an extension of Maxwell's electrodynamics in the framework of Einstein's special relativity, Mie's theory attempted to explain the elementary particles known at that time, electrons and protons, as the product of a universal electromagnetic field, thereby overcoming the conceptual duality between field and matter. Mie also tried to include gravitation in his electrodynamics, with the goal of a unifying theory

of physics, a precursor to the efforts of Hilbert and Einstein. Mie also corresponded with Einstein on the interpretation of gravitational waves. (pp. 9, 123–124, 147)

Millikan, Robert Andrews (1868–1953)

Photo: Harris & Ewing, AIP Emilio Segre Visual Archives, W. F. Meggers Gallery of Nobel Laureates

Robert Millikan was an American experimental physicist. After receiving his doctorate at Columbia University and studying in Berlin and Gottingen, Millikan joined the University of Chicago, where he became a full professor in 1910. He determined the charge of an electron and proved that all electrons carry the same charge, thereby demonstrating the atomic structure of electricity. He experimentally confirmed Einstein's photoelectric equation. In 1923, he received the Nobel Prize for these two achievements. His studies of the Brownian motion in gases put an end to all opposition to the atomic and kinetic theories of matter. Millikan was an eminent teacher and devoted time to producing textbooks and simplifying the teaching of physics. In 1921, he left the University of Chicago to become the director of the newly established Norman Bridge Laboratory of Physics at Caltech, where he conducted pioneering studies of cosmic radiation using ionization chambers. He invited Einstein to Caltech and tried for many years to bring him there permanently. (pp. 13, 62)

Noether, Emmy (1882–1935)

The German mathematician Emmy Noether is considered to be one of the founders of modern algebra, but she had to overcome prejudice and opposition to women's rights in order to study and pursue an academic career before she was able to obtain a professorship at the University of Göttingen in 1922. By that time, Noether had already made significant contributions to the theory of algebraic invariants, such as the so-called Noether's theorems, published in 1918, about the relation between symmetry and conservation laws in physics. They provide a general mathematical framework for the law of energy-momentum conservation in Einstein's general theory of relativity. In the following years, Noether contributed to topology and other fields of mathematics; above all, she created and developed the new field of abstract algebra. (pp. 7, 29–30)

Nordström, Gunnar (1881–1923)

The Finnish theoretical physicist Gunnar Nordström became a lecturer at the University of Helsinki in 1910 and later a professor at the technical university there. As a student, he had already written about Minkowski's electrodynamics and about the theory of relativity. In 1912, he formulated a relativistic theory of gravitation based on a scalar gravitational field. Nordström himself discussed this subject with Einstein on several occasions, for instance, in Zurich in 1913 and in Berlin in 1914. Later he published several articles on Einstein's general theory of relativity. They corresponded on various aspects

of the theory, specifically on the problem of gravitational waves. Nordström also solved the field equations for a spherically symmetric charged body. (pp. 8, 9, 47, 70, 99, 101)

Pauli, Wolfgang Ernst (1900–1958)

Photo: von Bettina Katzenstein / ETH Zürich (ETH-Bibliothek Zürich, Bildarchiv)

After receiving his doctoral degree in 1921 as a student of Sommerfeld at the University of Munich, Wolfgang Pauli spent time at the Universities of Göttingen, Copenhagen, and Hamburg before his appointment as professor of theoretical physics at the Federal Institute of Technology in Zurich. Pauli made groundbreaking contributions in modern physics, primarily in quantum mechanics. In 1945, he received the Nobel Prize in Physics for his "decisive contribution through his discovery of a new law of nature, the exclusion principle or Pauli principle." The discovery involved a relation between the spin of the electron and the atomic structure, which is the basis of a theory of matter. Pauli was the first to recognize the existence of the neutrino, an uncharged and massless particle, which emits energy as ß-radioactivity. He was one of the pioneers of the quantum theory of fields. In 1921, while still a student, he published a comprehensive review of the theory of relativity that became a seminal text during the formative years of general relativity. Already in the early stages of the quest for a unified field theory and contrary to Einstein, Pauli did not believe that such a theory, which would also account for the origin and nature of elementary particles, is possible in the framework of classical continuous fields. Nevertheless, he and Einstein corresponded extensively during Einstein's efforts in this direction. (pp. 9, 47, 133, 140, 147)

Poincaré, Henri (1854–1912)

Photo: Deutsches Museum, München, Archiv

In 1879, Henri Poincaré graduated from the prestigious École de Mines with a degree in mining engineering, and in the same year he obtained a doctoral degree from the University of Paris in mathematics. Although he did not abandon engineering, his most important achievements are in mathematics, mathematical physics, and celestial mechanics. His research on the three-body problem led to the discovery of deterministic chaos, which laid the foundations of modern chaos theory. He is also considered to be one of the founders of the field of topology. His work on mathematical features of the Lorentz transformation was an important precursor in the development of special relativity and generated a debate about the distribution of credit for the discovery of the theory of special relativity between Lorentz, Poincaré, and Einstein. Einstein did not acknowledge this part of Poincaré's work, but he did acknowledge his analysis of the delicate relation between fundamental concepts such as space, time, and experience. Poincaré as a philosopher of science and mathematics is best known for his conventionalism. Thus, although Poincaré recognized that Euclidean and non-Euclidean geometry are equally valid, he argued that our experience has and will continue to predispose us to formulate physics in terms of Euclidean geometry because of pure convention. (pp. 7, 26, 30, 32, 106, 125, 130, 163 [3], 266 [106])

Reichenbach, Hans (1891–1953)

Photo: Vienna Museum

Hans Reichenbach was born in Hamburg. He studied physics, mathematics, and philosophy in Berlin, Göttingen, and Munich, receiving his PhD in Erlangen with a thesis on the concept of probability. In the winter semester of 1917–1918, he continued his studies in Berlin, where he took a course on relativity with Einstein. Together with Moritz Schlick, he soon became one of the most prominent advocates and interpreters of relativity theory. He criticized popular misunderstandings of the theory and developed his own views on science in the context of philosophical discussions on relativity. Together with Rudolf Carnap, Herbert Feigl, Victor Kraft, and others, he became one of the founders of logical positivism. Following a proposal by Einstein, in 1926 Reichenbach became professor of philosophy of physics at the University of Berlin. He was dismissed in 1933 after the rise of the Nazi regime and had to emigrate. (pp. 4, 8, 9, 73, 111, 112–113, 114, 116)

Reissner, Hans (1874–1967)

Photo: © Getty Images

Hans Reissner had a distinguished career as a professor, researcher, and aeronautical engineer during the development of early twentieth-century German aviation. He designed the first successful aircraft with all-metal wing and tail surfaces and the first controllable pitch propeller. In addition to his professional career in aeronautical engineering, he pursued his deep interest in mathematical physics, specifically in the theory of relativity. He was the first to solve Einstein's equation for the metric of a charged point mass. Reissner immigrated to the United States in 1938 and taught at the Illinois Institute of Technology and the Polytechnic Institute of Brooklyn. (pp. 8, 47)

Schlick, Moritz (1882–1936)

The German philosopher Moritz Schlick was the founder and a leading figure of the Vienna Circle of logical empiricism. After completing his studies in physics, Schlick turned to philosophy and concerned himself with problems of ethics, epistemology, and philosophy of science. In 1922, he was appointed professor at the University of Vienna as a successor to Ludwig Boltzmann and Ernst Mach. Schlick published a paper in 1915 on the philosophical implications of Einstein's theory of relativity, pointing out Einstein's clarification of the concept of simultaneity at a distance. In a subsequent book about the concepts of space and time in contemporary physics, Schlick explained Einstein's adoption of non-Euclidean geometry on the basis of geometric conventionalism. In his major work on the theory of knowledge, which was published in 1918, Schlick argued that the physically real is characterized by spacetime coincidences, a concept that became instrumental in overcoming Einstein's infamous hole argument. (pp. 4, 8, 9, 107–9, 111, 113, 116, 117–119, 146)

Schwarzschild, Karl (1873–1916)

The German astrophysicist Karl Schwarzschild was professor of astronomy at the University of Göttingen and director of the local observatory from 1901 to 1909. Later, he became director of the astrophysical observatory in Potsdam. In his early years, his work focused on celestial mechanics and stellar photometry. He was the first to systematically use photography in astronomical observations. In later work, he also concerned himself with problems of geometrical optics, electrodynamics, and spectroscopy applied to astrophysical phenomena. In 1914, Schwarzschild attempted to observe a gravitational red shift in the solar spectrum, as predicted by Einstein's general theory of relativity. In 1916, he found the first exact solution of Einstein's field equations for a static, spherically symmetric mass, which led to an exact calculation of Mercury's perihelion motion. His correspondence with Einstein on gravitational waves was a trigger to Einstein's seminal treatment of this subject. (pp. 4, 8, 22, 46–48, 54, 94, 96–97, 98, 102, 155, 391)

Seeliger, Hugo von (1849–1924)

Photo: AIP Emilio Segre Visual Archives, Brittle Books Collection

Hugo von Seeliger was born in Bielsko-Biała, Austrian Silesia, and studied astronomy and mathematics in Heidelberg and Leipzig. After 1882, he became director of the observatory at the University of Munich. Karl Schwarzschild was one of his most prominent students. Seeliger mainly worked as a theoretician, focusing on celestial mechanics, stellar statistics, and the illumination of Saturn's rings. He also developed what was called the zodiacal light hypothesis. After the completion of general relativity, Seeliger became involved in a controversy with Erwin Freundlich about its astronomical verification. (pp. 54, 79)

Silberstein, Ludwik (1872–1948)

Photo: J. W. Perry, Hilger and Watts Ltd., courtesy AIP Emilio Segre Visual Archives

Ludwik Silberstein was a Polish-American physicist. After receiving a doctoral degree in Berlin, he taught at Universities of Lvov, Bologna, and Rome before moving to the United States. Between 1920 and 1930, he worked as a physicist for the Eastman Kodak Company and then as its consultant until the end of his life. In 1911, he published an alternative formulation of the special theory of relativity and in 1918 suggested a different (incorrect) version of general relativity. In 1933, he challenged the validity of Einstein's theory of general relativity, which resulted in a three-year-long intense controversy between the two scientists on the two-body problem in general relativity. In the early 1920s, Silberstein became a pioneer in delivering full courses on relativity, teaching at the Universities of Toronto and Chicago and Cornell University. (p. 150)

Sitter, Willem de (1872–1934)

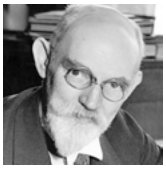


Photo: University of Chicago Photographic Archive, Special Collections Research Center, University of Chicago Library

In 1908, the Dutch astronomer Willem de Sitter was appointed professor at the University of Leiden, where he became director of the observatory in 1919. His main contributions concerned celestial mechanics, stellar photometry, the measurement of stellar parallaxes, and the application of the theory of relativity to cosmology. In 1913, de Sitter provided astronomical evidence for the constancy of the velocity of light, thus confirming Einstein's special theory of relativity. In the years 1916–1917, he published three papers on the astronomic consequences of Einstein's gravitational theory, which attracted the attention of Arthur Stanley Eddington and led to his solar eclipse expedition in 1919. During the 1920s, Einstein and de Sitter engaged in a scholarly debate on the validity of their cosmological models. In 1932, they collaborated on a paper about the expansion of the universe, in which they argue for the probable existence of matter that does not emit light. This paper constituted the standard cosmological model for many years to come. (pp. 4, 8, 9, 22, 48, 49, 62, 66, 69–72, 75–78, 79, 82, 85, 88–89, 97–98, 109, 144, 147, 155, 391, 392, 394)

Slipher, Vesto Melvin (1875–1969)

Photo: The Special Collections Research Center, University of Chicago Library

Slipher was an American astronomer who spent his entire career at the Lowell Observatory in Flagstaff, Arizona. He investigated the rotation period of planets, the composition of planetary atmospheres, and discovered the rotation of spiral galaxies. In the context of the present book, he is best known for his measurements, published in 1917, of the redshifts of 25 distant nebulae, later identified as galaxies. Thus, he was the first to discover galactic redshifts, which were identified as Doppler shifts. This discovery, combined with de Sitter's model, could have led to the idea that the universe was expanding, but apparently it was too early for such a revolutionary conclusion. Slipher preceded, by a whole decade, Hubble's work, which combined such redshift measurements with a reliable estimate of distance to the galaxies and provided conclusive evidence about the expansion of the universe. (pp. 49, 82, 85)

St. John, Charles Edward (1857–1935)

After studying at the University of Michigan and Harvard University, and after teaching at Michigan, Charles St. John began teaching astronomy at Oberlin College, in Ohio, in 1897, spending his summers at Yerkes Observatory and improving the astronomy facilities at Oberlin. In 1908, he was invited by George Hale to develop the solar research program at Mt. Wilson Observatory. St. John's research there focused on different

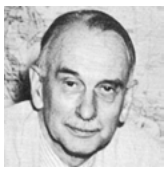
aspects of solar astronomy: sunspots, the structure of the solar atmosphere, and use of spectroscopic methods to examine its chemical composition. He also made observations of the planet Venus, showing that there was insufficient oxygen in the atmosphere to support life as we know it. In his later work, he spent much time measuring the gravitational red shift of solar spectral lines, attempting to confirm Einstein's theory of general relativity. (pp. 60, 64, 65)

Thirring, Hans (1888–1976)

Photo: ÖNB/Wien

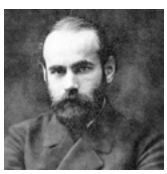
Between 1910 and 1938, Hans Thirring served as a professor at the Institute of Theoretical Physics at the University of Vienna, and beginning in 1921, as its president. He suggested to his colleague Josef Lense at the university that using perturbation theory, they try to derive from Einstein's gravitational field equation the equation of motion of a particle in a field generated by a rotating mass, with application to the orbits of planets and the moon. This resulted in a joint paper and in the discovery of the Lense-Thirring frame-dragging effect. This paper and his book T*he Idea of the Theory of Relativity* (1921), which was translated into English and French, brought him international acclaim. After the annexation of Austria by Nazi Germany in 1938, he was forced to retire, mainly because of his pacifist views and his endorsement of the theory of relativity. After WWII, he was reinstated at the university and became politically active, representing the Austrian Socialist Party in the parliament. (pp. 80–81, 117, 148–149, 262 [102], 398)

Tolman, Richard Chace (1881–1948)



Richard Tolman was an American physicist and physical chemist who specialized in thermodynamics and statistical mechanics. In 1922, he was elected a fellow of the American Academy of Arts and Sciences and in the same year joined the faculty at Caltech, where he became a professor of physical chemistry and mathematical physics and, subsequently, dean of the graduate school. In the later years of his career, Tolman became interested in the application of thermodynamics to relativistic systems and cosmology. In 1934, he published the influential monograph *Relativity, Thermodynamics, and Cosmology*, which contained important results on black body radiation in an expanding universe that later became relevant for the understanding of the cosmic microwave background radiation. During World War II he was vice-chairman of the National Defense Research Committee and chief science adviser to the head of military affairs concerning the development of the atomic bomb. (pp. 49, 63, 85, 87, 156)

Wertheimer, Max (1880–1943)



Max Wertheimer was born in Prague and after completing his PhD in 1904 in Würzburg, worked as a psychologist in Prague, Frankfurt, and Vienna. In Frankfurt, he laid the foundation of Gestalt psychology, together with his colleagues Wolfgang Köhler and Kurt Koffka. Psychology at that time emancipated itself from philosophy and developed its own experimental paradigms. But in contrast to

empiricist approaches and their focus on the description of human behavior, Wertheimer was interested in thought processes, which he tried to capture not just by introspection but also with his own experimental methodology, which contrasted with the traditional, more philosophical approach. From 1916 to 1929, he was active at the University of Berlin. In his most famous book, *Productive Thinking*, which was published posthumously in 1945, he reported on his intense discussions with Einstein, beginning in 1916, in which he tried to unravel the thought processes that had led to special relativity. Only in 1929 did Wertheimer become an ordinary professor at the University of Frankfurt. In 1933 he was forced by the Nazi regime to leave Germany because he was Jewish; he emigrated to the United States, where he became a professor of philosophy at the New School for Social Research in New York. (p. 118)

Weyl, Hermann (1885–1955)

Photo: ETH-Bibliothek Zürich, Bildarchiv

In 1913, the German mathematician Hermann Weyl was appointed professor at the ETH Zurich, where he met Einstein and started to work on mathematical aspects of general relativity. In 1918, Weyl published a landmark book on relativity, *Space, Time, Matter*, in which he further developed and applied the affine connection formulation of differential geometry. In that book he also formulated the first attempt to unify gravitation and electromagnetism by extending Riemannian geometry. In 1930, Weyl took over David Hilbert's position in Göttingen but then immigrated to the United States in 1933, joining Einstein at the Institute for Advanced Study in Princeton, New Jersey. Weyl made important contributions to the development of theoretical physics. His main concern in the early years had been analysis and spectral theory, but he later became interested in topology and differential geometry. In the following years, he worked on the theory of groups and its applications to quantum mechanics. In addition, Weyl concerned himself throughout his scientific career with the foundations of mathematics and philosophy of science. (pp. 4, 8, 9, 19, 20, 22, 31, 47–49, 77–78, 102, 116, 117, 125–131, 143–147, 229 [69], 232 [72], 254 [94], 258 [98], 302 [142], 399)

INDEX

The entries in square brackets refer to the original pagination of Einstein's book. His popular lectures are not indexed. Bold entries refer to key terms.

A
Abraham, Max, 94–95, 96, 100, 102, **389**
acceleration, 34, 35, 38, 41, 55, 95 179 [19], 216 [56], 217 [57], 260 [100]
  inductive action of, 262 [102]
action, 29, 128
  at a distance, 39, 94, 115, 122
  principle, 126
  and reaction, 95
Adams, Edwin P., xi, 5, 17, 21, 79, 125, 363
Adams, Walter Sydney, 61, 64–66, **389–390**
Adler, Friedrich, 73
affine connection, 19, **20**, 126, 129, 132, 144, 407
age of universe. *See* universe
Akademie Olympia, 26, 106
anti-relativity campaign (also literature), 9, **150ff.**
anti-Semitism, 9, 12, 62, 117, 140, 152, 397, 398
approximation
  methods, 22, 46, 96, 97, 241 [81]
  solution, 50, 96, 98, 101, 241 [81] ff.
astronomy, 52, 54, 59, 62, 75, 141, 150, 248 [88], 258 [98], 290 [130]
  extragalactic, 82
  radio, 103
  solar, 65
  theoretical, 257 [97]
atoms, 61, 99, 114, 123, 124, 126, 207 [47]
  atomic character of energy, 137
  atomic nature of electricity, 147
  atomistic nature of physics, 137, 145
axiomatic geometry. *See* geometry

B
Bachem, Albert 60
Ben-Menachem, Yemima, xiv, 26
bending of light. *See* light
Bergson, Henri 26, 109
  his concept of time, 26, 109
Bern conference, 103, 157
Besso, Michele, 53, 61, 75, 80, 86, 95, 106, 142, **390**
Bianchi identities, 302 [142] ff., 347 [154]
Big Bang, 51, 90, 398
Biot-Savart force, 202 [42]
binary system, rotating 100, 101, 103
black holes, 46, 50, 96, 101, 102, 103, 104
Blumenfeld, Kurt, 11
Born, Max, 9, 94, 102, 149, 390
boundary conditions, 36, 43, 69, 70–71, 73, 79, 81, 263 [103], 264 [104], 267 [107], 357 [164], 358 [165]
British Royal Astronomical Society, 56
Budde, Arnold Emil, 58
bucket experiment. *See* Newton

C
California Institute of Technology (Caltech), xiv, 13, 62–63, 66, 85, 401, 406
Campbell, William Wallace 57, 64, 65, **390**
Cassirer, Ernst 8, 9, 113, 146, **391**
centrifugal
  field, 42, 260 [100]
  force, 32, 34, 42, 79, 80, 220 [60], 262 [102]
Chapel Hill conference, 103, 157
charged particles, origin of, 147
Chicago, University of, 13, 62, 389, 401, 404
Christoffel, Elwin, 19
City College of New York, 13
clock(s), 31, 42, 43, 53, 62, 127, 161–162 [1–2], 176 [16], 187–188 [27–28], 205 [45], 220 [60], 224 [64], 252 [92], 288 [128]

clock(s) (*Continued*)
moving, 185 [25], 196 [36]
rotating, 220 [60]
clocks and rods, 28, 31, 116, 122, 126, 127, **196 [36]**, 220 [60], 223 [63], 250 [90]
Cohen, Morris, 13
Columbia University, 13, 401
congruence, 163 [3]
conservation law(s) (principles), 29, 30, 43, 210–211 [50–51], 214 [54], 220 [60], 347 [154]
of electric charge current, 136
of energy momentum, 8, 18, 21, 29, 30, 42, 72, 123, 124, 144, 243 [83], 356 [163]
continuum
Euclidean, 229 [69], 235 [75], 237 [77]
field theory, 138, 147, 264 [104]
four dimensional, 18, **28**, 144, 190–191 [30–31], 199 [39], 223–224 [63–64], 239 [79], 265 [105]
non-Euclidean, **30**
relativistic theory, 28, 145
spacetime. *See* spacetime
three-dimensional, 264 [104], 265 [105]
conventionalism, 106, 110, 115, 116, 402, 403
coordinates, 19, 106
admissible, 42, 43, 97
Cartesian, 18, 26–27, 30, 164 [4] ff., 222 [62]
choice of, **274 [114]**
conditions, 46, 96, 97, 99
curvilinear, 100, 221 [61]
de Sitter's, 100
Finkelstein, 48
Galilean, inertial, 35, 186 [26]
meaning of (role), 30, 102, **107**
polar, 46, 256 [96]
preferred system of, 27, 167 [7]
restrictions on admissible, 42
rotating, 80
transformation, 27, 167 [7], 186 [26], 190 [30], 196 [36], 202 [42], 317 [157], 335 [142]
unimodular, 99
Coriolis force, 32, 42, 79, 80
cosmic radiation, 58
cosmology, 8, 49, 50, 62, 63, 65, 69, 156, 394, 397, 398, 405, 406
and exact solutions, **49**
history of, 62
observational, 4, 69, 156
relativistic, 8, 44, **69ff.,** 393
cosmological
consequences, 3, 4, 7, 8, 49, 62, 144, 392
problem, 8, 17, 22, 69, **78–79**, 81–82, 86, 89, 124, 258 [98], **269 [109] ff.**
cosmological constant, 48, 49, 72–73, 75, 81, 82, 86–90, 124, 156, 271 [111], 397
cosmological models (solutions), 3, 8, 49, **50**, 71, 78, 147
de Sitter's, 50, 75–76, 78, 82, 89, 147
Einstein's, 71, 75–76, 147
Einstein–de Sitter, 88–89
Friedmann, 49, 90. *See also* Friedmann
Lemaître's. *See* Lemaître
covariance, 171 [11] ff.
of equations of continuity, 180 [20]
general, 8, 30, 37, 40, 42, 43, 72, 317 [157], 331 [138], 347 [154], 397
of the laws of physics, 27
of pre-relativistic physics, 28
curvature (curved spacetime), 19, 49, 53, 73, 81–83, 86, 88–90, 102, 252 [92], 263–264 [103–104], 267 [107], 273 [113], 275 [115], 279–280 [119–120], **281 [121] ff.**

D

Dällenbach, Walter, 59
Davidson, Charles, 56–57
de Jans, Carlo, 49
Descartes, Rene, 107, 118
differential calculus, absolute, 19, 143, 295 [135] ff., 399
Doppler effect, 82, 86, 288 [128], 399, 405
Droste, Johannes, 8, 47, 49, **391**
dualism in physical theories, 115, 124
Duhem, Pierre Maurice Marie, 114, **391–392**
Duhring, Eugen, 106
Dyson, Frank W., Sir, 57

E

Earman, John, 60
eclipse. *See* solar
Eddington, Arthur Stanley, 8–9, 48, 56–57, 60, 62, 72, 101–102, 117, 125, 127, 128–130, **144–145**, 155, 390, **392**, 404
Ehrenfest, Paul, 11, 47, 59, 70–71, 108, 127, **392**
Einstein–de Sitter–Weyl–Klein debate, 8, 77–78
Eisenhart, Luther, 11, 15
Eisenstaedt, Jean, 4, 47, 50, 155–156

elasticity theory, 145
electrodynamics, 39, 95, 98, 99, 106, 136, 209 [49], 247 [87]
electromagnetic
energy, 99, 242 [82]
energy tensor, 124, 207 [47], 210 [50]
field, 29, 35, 78, 91, 98, 122, 123–124, 200 [40], 201 [41], 204 [44], 247 [87], 257 [97], 258 [98], 289 [129], 293 [133]
field tensor, 257 [97], 313 [153]
potential, 98, 122, 126, 130
radiation, 95, 101
retarded potentials, **98**
waves, 91, 95, 98, 122
electron, 47, 123, 124, 128, 130, 131, 145, 209 [49]
theory, 95, 181 [21], 207 [47]
electron-proton mass difference, 129, 130
elementary particles, 123, 124, 128, 129, 147
Enebo, Sigurd, 58,
energy and mass, 38, 145, 203 [43], 207 [47], 211 [51]
energy-momentum. *See* conservation laws
Engler, Fynn Ole, 107
Entwurf theory, 29, **42–43**, 46, 47, 53, 80, 94, 107, 390
epistemology, 8, 17, 25, 36, 69, 75, 91, 106–108, 111, 115, 116, 118–119, 136, 144, 268 [108]
genetic, 118
equation of continuity, 180 [20], 214 [54]
equivalence principle, 4, 19, 35–36, 38, 39, 44, 53, 124, 136, 333 [140]
ether, 7, 13, 39, 115–116, 142, 150, 186, 399
Euclidean. *See* geometry; space
Euler equations, 18, 213–214 [53–54]
Euler-Lagrange equations, 29
exact solutions, 8, 22, 46, 47, **49ff**., 147
expanding universe. *See* universe, expanding

F
field concept, 137, 322 [162] 325 [165], 333 [140]
field equations. *See* gravitational field equations; Maxwell's equations (theory); unified field theory
five-dimensional spacetime, 130–131
Fizeau, Hippolyte, 187 [27]
Flamm, Ludwig, 48
Fleck, Ludwig, 118–119
Freundlich, Erwin Finlay, 8, 54–55, 57–58, 59–60, 111, 141, **392–393**, 404
Friedmann, Alexander, 8, 50, 82–83, 85, 86, 90, 272 [112], 287 [127], **393**, 394, 398
Friedman, Michael, 106

G
galaxies (distant nebulae), 8, 49, 58, 63–65, 75, 82–84, 86, 90–91, 103, 156, 394, 405
Galilean regions, 219 [59]
Gamow, George, 87
gauge, 126, 132
field theories, **128**
invariance, 128
symmetry, 128
Gauss, Carl-Friedrich, 221 [61], 224 [64], 334 [141]
Gehrcke, Ernst, 151, 152
general relativity
golden age of, 4, 157
low-water-mark period, 4, 50, 91, 103, 155
Newtonian limit of, 22, 97–98
renaissance of, 4, 59, 103, 104, 153, 156–157
geodesic line, 20, 53, 131, 222 [162], 237 [177], 269 [109], 272 [112], 278 [118]
geometry, 18, 73, 77, 116, 127, 143
axiomatic, 25, 116
differential, 19
Euclidean, 13, 18, 26, 27, 30, 116, 126, 164 [4] ff.
non-Euclidean, 19, **30**, 122, 146, 221 [61]
and physics, 7, 15, **25**
practical, 25, 31, 114, 116
Riemannian, 122, 126, 129, 130, 143, 146
of spacetime, 19, 30, 122, 130
Gestalt psychology, 118
Global Positioning System (GPS), 62
Glymour, Clark, 60
Goenner, Hubert, 152
gravitational field equations, 8, 20, 22, 36–37, 39, 42, 46, 58, 69–73, 75, 77–79, 80, 82–83, 88, 90–91, 94, 96, 101, 107–108, 110, 123, 225, 129, 136, 141, 144, 208[48], 241 [81], 249 [89], 254–255 [94–95], 258 [98], 261 [101], 264–266 [104–106], 272 [110], 274 [114], **277 [117]**, 294 [134], **300 [140] ff.**, 307 [147], 309 [149]
gravitational
lensing, **57ff.**
light deflection, 3, 8, 11, 22, 47, 53ff., 59, 101, 110, 140, 391, 142, 144, 146, 155
redshift, 8, 15, 22, 49, 53, 59, 62, 63, 82, 252 [92], 253 [93], 264, 269 [109]
gravitational waves, 3, 8, 94, 96, 98, 100, 102, 155
dipole moment, 95, 96, 100, 101
quadrupole moment, 8, 95, 100, **101**, 102
speed propagation, 101

Gravity Probe B experiment, 81
Grebe, Leonhard, 60,
Grommer, Jakob, x, 70–71, 131, **393**
Grossmann, Marcel, 52, 53, 391, 399

H
Haber, Fritz, 12
Habicht, Conrad, 53, 106
Hadamard, Jacques, 48
Hale, George Ellery, 55, **393–394,** 395, 405
Hamilton's principle. *See* variational calculus (principle)
Hebrew University of Jerusalem, xiv, 11–12, **131**
Heckmann, Otto Hermann Leopold, 89, **394**
Hegel, Georg Wilhelm Friedrich, 118
Hentschel, Klaus, 60, 112, 115, 116
heuristics, 4, 22, 36–37, 38–40, 44, 69, 109–110, 124, 181 [21], 400
Hibben, John, ix, 11, 13, 363
Hilbert, David, 4, 9, 25, 29, 47, 48, 123, 393, **394**, 396, 401, 407
holism, 116, 392
Hubble, Edwin, 8, 63–65, 84–87, 88, 89–91, 272 [112], 278 [118] ff., **394–395**, 398, 405
  constant (time), 84, 86, 91
Hulse, Russell, 103
Humason, Milton Lasell, 63, 65, 85, **395**
Hume, David, 106
Huygens`s principle, 53
hydrodynamics, 18, 145, 211 [51]
hypothetico-deductive method, 114

I
induction and deduction, 114
Institute for Advanced Study, Princeton, 67
invariants (theory of), 29, 70, 122, 143, 169–170 [9–10], 172 [12], 173 [13], 175–177 [15–17], 181 [21], 192 [32], **197 [37] ff.**, 204–205 [44–45], 211–212 [51–52], 223–224 [63–64], 226 [66], 228 [68], 230–231 [70–71], 239 [79], 286 [126], 324 [164], 348 [155], 349 [156], 394, 401
isotropy of space. *See* space, isotropy of

J
Janssen, Michel, 41, 100
Jupiter, 59, 81

K
Kaluza, Theodor Franz Eduard, 9, 125–126, 130–132, 258 [98], **395**
Kaluza-Klein theory, 131–132
Kant, Immanuel, 26, 30, 106, 111–113, 115, 391
Kaufman, Brurya, 135
Klein, Felix Christian, 8–9, 29, 77–78, 392, **395–396**
Klein, Oskar Benjamin, 9, 131, 132, 394, 395, **396**
Kopff, August, 149, **396**
Kragh, Helge, 85, 89
Kretschmann, Erich Justus, 8, 9, 37–41, **39, 397**
Kuhn, Thomas, 4

L
Lagrangian. *See* variational calculus (principle)
Lanczos, Cornelius, 48, 49, 397
Laser Interferometer Gravitational Wave Observatory (LIGO), **102–104**
von Laue, Max, 9, 28, 48–49, 117, 145–146, **397–398**
Lawson, Robert, 115
Leibniz's prestabilized harmony, 110
Lemaître, Georges, 8, 48, 50, 62–63, 66, 83–85, 90, 395, **398**
  Lemaître-Tolman solution, 48
  primeval atom, 90
Lenard, Philipp, 152, **398**
Lense, Josef, 80, 81, 148, **398–399**, 406
Lense-Thirring effect, 80
  frame dragging, 148
Levi–Civita, Tullio, 19, 20, 47, 126, 129, 144, 224 [64], 302 [142], 334 [141], **399**
Lick Observatory, 57
Lie's theory of differential equations, 29
light
  bending of (deflection of). *See* gravitational light deflection
  cone, 199 [39]
  corpuscular nature of, 62
  propagation of, 54, 55, 187 [27] ff.
  ray, 186 [26], 189 [29], 192 [32], 252 [92]
  redshift. *See* gravitational redshift
  signal, 198 [38] ff.
  velocity of, 18, 94, 98, 101, 102, 187 [27] ff.
  waves, 142
limit of weak static gravitational fields, 42
line element, 46, 50, 73, 122, 126, 127, 129, 224 [64], 238 [78], 275 [115]
logical
  economy, 90
  empiricism, 111
Lorentz, Hendrik Antoon, 4, 9, 12, 18–19, 28–29, 31, 39, 47, 106, 145, 189 [29], 391, 392, **399**
Los Angeles Times, 63

M
Mach, Ernst, 8, 22, 34–36, 37, 38–40, 41, 42, 44, 49, 55, 69, 71, 75, 78, 79, 80, 81, 82, 106, 110, 111, 113, 115–116, 118, 216 [56], 259 [99], 260 [100], 263 [103], 267 [107], 268 [108], 333 [140], **399–400,** 403
Mandl, Rudi W., 58–59
magneto-electric induction, 35
Manifesto of the 93, 14
Manifesto to the Europeans, 14
mass and energy. *See* energy and mass
Maxwell's equations (theory), 7, 18, 22, 28, 29, 98, 122, 123, 126, 136, 144, 148, 181 [21], 200 [40], 209–210 [49–50], 257 [97], 266–267 [107–108], 294 [134], 307 [147], **312 [152]**, 315 [155], **328 [135]**, 332 [139], 352 [159]
Mayer, Walther, 62, 64, **400**
mechanics
  classical, 18, 29, 34, 36, 39, 78, 80, 82, 95, 106, 142, 180 [20], 184–186 [24–26], 190 [30], 206–207 [46–47], 211 [51], 214 [54], 216 [56], 218 [58], 256 [96]
  history of, 107
  quantum, 4, 91, 117, 128, 132, 137, 321–322 [161–162], 324 [164]
  statistical, 85
Mercury's perihelion motion, 22, 46, 47, 52–53, 60, 80, 96, 97, 110, 254 [94], 257 [97]
Methuen Verlag, xi, xii, 5, 135
metric tensor, 19, 20, 30, 37, 39, 70, 72, 81, 82, 97, 122, 123, 126, 127, 129, 132, 224 [64] ff., 227 [67] ff., 295 [135], **299** [**139**] **ff.**
Meyerson, Émile, 118, **400**
Michelson, Albert, 13, 65, 365, 370
Michelson and Morley experiments, 187 [27]
Mie, Gustav, 9, 123–124, 147, **400**
microwave background radiation, 89
Millikan, Robert, 13, 62, **401**
Milky Way, 70, 75, 82
Minkowski, Hermann, 18, 28–29, 80, 95, 182 [22], 191 [31], 192 [32], 198 [38], 199 [39], 401
monopole and dipole radiation, 100. *See also* gravitational waves
Mount Wilson Observatory, 55, 60, 61, **62ff.**, 84, 86, 88

N
nebulae. *See* galaxies
neo-Kantian perspective, 111, 113
New York Times. See *The New York Times*
Newton
  concept of absolute space, 110, 115, 398
  constant, 249 [89]
  law of gravitation, 4, 38, 52, 79, 148
  mechanics. *See* mechanics, classical
  particle theory of light, 53
  rotating bucket experiment, 34, 35, 40, 42–43, 80
Nicolai, Georg, 14
Noether, Emmy, 7, 29, 30, **401**
non-Euclidean. *See* geometry
non-symmetric field. *See* unified field theory
Nordström, Gunnar, 8–9, 47, 70, 99, 101, **401–402**

P
parallel transport (displacement). *See* vector, parallel transport of
Pauli, Wolfgang Ernst, 9, 47, 133, 140, 147, **402**
perihelion. *See* Mercury's perihelion motion
photoelectric effect, 62
Piaget, Jean, 118
Planck, Max, 107, 110, 397
Poincaré, Henri, 7, 26, 30, 32, 106, 125, 130, 163 [3], 266 [106], **402**
point coincidence, 37, 38, 107, 108
Poisson's equation, 20, 22, 242–244 [83–84]
Pollack, Leo Wentzel, 54
Popper, Karl, 114
positivism, 110, 115, 118
Potsdam Observatory, 58
pre-relativity physics, 17–18, 185 [25]
primordial nucleosynthesis, 89
Princeton University, x–xiii, 5, 11, 14, 15
principle of inertia, criticism of, 215 [55], 218 [58], 239 [79]
psychology, 15, 106, 115, 118

Q
quadrupole. *See also* gravitational waves
  formula, 8, 102
  moment, 95, 100
  radiation, 101
quantum mechanics. *See* mechanics
quantum theory of gravitation, 99
quasars, 4, 59, 157

R
redshift. *See* gravitational redshift
reference
  accelerated frame of, 34, 35, 55, 69
  bodies of, 26, 27, 28, 163 [3]
  de Sitter's frame of, 98
  frame of (space of), 18, 19, 22, 26, 27, 38, 41, 42, 99, 142, 163 [3] ff., 176 [16]
  inertial frame of, 18, 28, 35, 129
  rotating frame of, 31, 42, 80, 100, 101
Reichenbach, Hans, 4, 8, 9, 73, 111–112, 113–114, 116, **403**
Reissner, Hans, 8, 47, **403**

Reissner-Nordström metric, 47
relativistic continuum theory, 28, 145
relativity
distinction between kinematical and physical, 41
of the gravitational field, 43
relativity principle, 4, 7, 18–19, **27**–**28**, **34–**38, **40**, 70, 107, 110
Ricci–Curbastro, Gregorio, 19, 399
Richter-Bozen, Gustav, 151
Riemann, Bernhard, 15, 28, 136, 141, 224 [64]
Robertson, Howard Percy, 49
rods, measuring (rigid), 26, 30ff., 103, 151, 175 [15], 196 [36], 219 [59], 229 [60], 222 [62], 223 [63], 251 [91]
rods and clocks. *See* clocks and rods
Rosen, Nathan, 101
Rosenberger, Ferdinand, 106
rotating
bodies, 34, 40, 41, 42, 43, 79, 80, 101, 260 [100], 262 [102]
bucket, 34, 80
clocks, 220 [60]
coordinates. *See* coordinates
disk, 15, 19, **31**, 32
Royaumont conference, 157
Russell, Bertrand, 117,
Ryckmann, Thomas, 106

S
Sandage, Allan, 91
Sauer, Tilman, 58, 132
Schlick, Moritz 4, 8–9, 107–109, 111, 113, 116, 117–119, 146, **403**
Schrödinger, Erwin, 4, 117, 132–133
Schwarzschild, Karl, 4, 8, 22, 46–47, 48–49, 50, 54, 94, **96–97**, 98, 102, 155, 254 [94], 391, **404**
radius, 47, 48, 50
solution, 22, **46–50**, 255 [95]
Seeliger, Hugo von, 54, 79, **404**
Silberstein, Ludwik, 150, **404**
simultaneity, 176 [16], 187 [27] 190 [30], 322 [162]
singularities, 48, 49, 50, 77, 78, 90–91, 128, 130, 280 [120], 284 [124], 289 [129]
free solutions, 130, 131, 138, 325 [165]
Sirius, 61, 269 [109]
de Sitter, Willem, 4, 8–9, 22, 48–49, 50, 62, 66, 69–71, 72, 75–78, 79, 82, 85, 88–89, 97–98, 100, 109, 144, 147, 155, 187 [27], 391, 392, 394, **405**
skew-symmetrical tensor, 174 [14], 175 [15], 178 [18], 179 [19], 183 [23], 200 [40], 233 [73], 234 [74], 237 [77], 294 [134], 297 [137]
Slipher, Vesto Melvin, 49, 82, 85, **405**
Smith, Robert W., 85
solar
Australian eclipse (1922), 57
British expedition (1919), 8, 56, 101, 110, 253 [93], 396
eclipse, 3, 54, 55, 57
German expedition (1914), 56, 146
spectrum, 60, 393
Solovine, Maurice, x, xi, 26, 106, 133–134, 148
Solvay conference, 11, 12, 84
Sommerfeld, Arnold, 52, 59, 96, 140, 147, **402**
space
absolute, 34, 35, 38, 69–70, 110, 115, 151
concept (perception) of, 13, 14, 26, 39, 78, 106, 111, 163 [3]
empty, 270 [110], 293 [133], 307 [147], 324 [164], 325 [165], 328 [135], 330 [137], 332 [139]
Euclidean, 13, 26, 28, 48, 73, 199 [39], 276 [116], 334 [141]
flat, 100
four-dimensional, 95, 130, 132, 198 [38], **272 [112] ff.**
homogeneity of, 176 [16]
isotropy of, 176 [16], 262 [112] ff.
three-dimensional, 30, 89, 164 [4], 178–179 [18–19], 191 [31], 199 [39], **272 [112] ff.**
symmetry of. *See* symmetry
spaces of reference, 18, 26, 165 [5], 184 [24], 186 [26], 215 [55], 216 [56]
Spaulding, E. G., 15
special relativity, 4, 7, 17, **18**, 28, 31, 35, 97, 103, 106, 118, 122, 124, 156, 157, 175 [15], 182 [22], **184–214 [24–54]**, 266 [66], 285 [125]
spectral lines, 15, 22, 53, 59, 61–62, 82, 90, 127, 223 [63], 252 [92], 269 [109], 272 [112], 288 [128]. *See also* gravitational redshift
St. John, Charles Edward, 60, 64, 65, **405**
string theory, **132**
structure of electrons, 123, 124, 130, 145
von Struve, Georg Otto Hermann, 54, 60
symmetry, 7, 18, 26, 28, 29, 30, 98, 100, 128, 174 [14], 182 [22], 183 [23], 200 [40], 202 [42], 226 [66], 227[67], 231 [71], 237 [77], 254 [94], 273 [113], 274 [114], 296 [136], 338 [145], 341 [148]

T
Taylor, Joseph, 103
tensor, 27, 136, 171 [11] ff., 199 [39] ff., 225 [65] ff., 242 [82], 296 [136] ff., 336 [143]

addition and subtraction of, 173 [13]
calculus, 18, 19, 146, 184 [24]
Christoffel, 19
contraction of, 174 [14]
covariant, 295 [135], 347 [154], 352 [159]
curvature, **338** [**145**] ff.
densities, 298 [138] ff., 352 [159]
differentiation of, 228 [68] ff., 232 [72] ff.
energy-momentum (stress energy), 18, 21, 39, 40, 87, 129, 146, 181 [21], 207 [47] ff., 211 [51] ff.
energy tensor of matter, 38, 211 [51], 243 [83] ff., 247 [87], 277 [177]
Kronecker, 295 [135], 298 [138], 341 [148]
metric (fundamental). *See* metric tensor
pseudo, **343** [**150**] **ff**.
Ricci, 129, 136
Riemann, 19, 20, 129, 136, **235** [**75**] **ff**., 237 [77], 244 [84], 259 [99]
special, 175 [15] ff.
stress, 181 [21]
symmetric, 174 [14], 180 [20] ff., 210 [50], 293 [133] ff., 328 [135], 334 [141], 342 [149]
theory of principle, 114
thermodynamics, 85, 106, 114, 156
Thirring, Hans, 80–81, 117, 148–149, 262 [102], 398, **406**
Thorne, Kip, 4, 157
thought experiments, 31, 35, 40, 42, 44, 48
Tolman, Richard Chace, 49, 63, 85, 87, 156, **406**
The New York Times, 5, 12, 21, 30, 63, 64, 79, 125, 133, 134

U
unified field theory, 9, 40, 63, 65, 82, 89, 101, 111, 115, **122ff.**, 131, 134, 137, 147
non-symmetric, 9, 132–133, 135–136
universe, 14, 22, 34, 39, 41, 63, 69ff., 81, 104, 156, 162 [2], 259 [99], 263 [103], 267 [107]
age of, 84, 86, **91**, 280 [120], 289 [129], 290 [130], 292 [132]
curvature of, 82, 267 [107]
density of, 90, 268 [108]
de Sitter's, 49, 50, 76–78
Einstein's, 22, 50, 64–65, 71ff., 76
Einstein-De Sitter, 88ff.
expanding, 8, 49, 50, 65, 82ff., 290 [130]
finite, 13, 250 [90], 259 [99], 270 [110] ff.
Friedmann's, 90, 272 [112] ff.,
geometry of, 29–30
model of a spherical, static, and closed, 22, 65, 71ff., 76
radius of, 77, 82, 86, 90, 267 [107], 276 [116]
University of Chicago, 13, 62

V
variational calculus (principle), 5, 29, 47, 129, 300–303 [140–143], 305 [145], 308–309 [148–149], **347** [**154**] **ff**., 355 [162]
Hamiltonian (principle), **29**, 136, **303 [143]**, 309 [149]
Lagrangian, **29**
vector, 18, 27, 29, 137, 144, 170 [10] ff., 177–180 [17–20], 182 [22], 202 [42] ff., 230 [70], 299 [139]
contravariant, 225–228 [65–68], 229 [69], 336 [143]
covariant, 225–227 [65–67], 233 [73], 324 [164]
parallel transport (displacement) of, 19, 20, 126, 129, 144, 295 [135], 300 [140], 336 [143], 339 [146]
tangent, 20
velocities, addition theorem of, 197 [37]
le Verrier, Urbain, 52, 257 [97]
Vienna Circle, 403
Vieweg Verlag, x, xi, 5
viscous compressible fluid, 180 [20]
Vogtherr, Karl, 151

W
Wazeck, Milena, 150
Weber, Joseph 103
Weimar Republic, 117, 152
Weinmann, Rudolf, 151–152
Weizmann, Chaim, 12
Wertheimer, Max, 118, **406–407**
West, Andrew, 13–14
Weyl, Hermann 4, 8–9, 19, 20, 22, 31, 47, 48, 49, 77–78, 102, 116, 117, 125–128, 129–131, 143–147, 229 [69], 232 [72], 254 [94], 258 [98], 302 [142], 399, **407**
Weyl's unified theory, 127, 145, 147
Will, Clifford, 4, 156,
Wittgenstein, Ludwig, 117
World War I, 11, 14, 56, 96
World War II, 4, 103

Z
Zangger, Heinrich, 58, 109, 127–128
Zionist delegation, 11–12